University of Alberta

Library Release Form

Name of Author: *Jacqueline Paulina Leighton*

Title of Thesis: *An Alternate Approach to Understanding Formal Reasoning: Thinking According to the Inductive-Coherence Model*

Degree: *Doctor of Philosophy*

Year this Degree Granted: *1999*

University of Alberta

An Alternative Approach to Understanding Personal Autonomy

[illegible]

A thesis submitted [illegible]

[illegible]

Department of Philosophy

Edmonton, Alberta

[illegible]

But certainly, for us who understand life, figures are a matter of indifference. I should have liked to begin this story in the fashion of the fairy-tales. I should have liked to say: "Once upon a time there was a little prince who lived on a planet that was scarcely any bigger than himself, and who had need of a sheep..."

To those who understand life, that would have given a much greater air of truth to my story.

-- Antoine de Saint Exupéry

The Little Prince, 1943

University of Alberta

An Alternate Approach to Understanding Formal Reasoning:

Thinking According to the Inductive-Coherence Model

by

Jacqueline Paulina Leighton ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of *Doctor of Philosophy*

Department of *Psychology*

Edmonton, Alberta

Fall 1999

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled *An Alternate Approach to Understanding Formal Reasoning: Thinking According to the Inductive-Coherence Model* submitted by *Jacqueline Paulina Leighton* in partial fulfillment of the requirements for the degree of *Doctor of Philosophy*.

To Greg

Abstract

The purpose of this paper is to explore an alternate account of participants' performance on formal reasoning tasks. It is proposed that participants approach formal tasks inductively because our reasoning system is best characterized as consisting of inductive underlying processes, and errors made on formal tasks are best understood applying this inductive hypothesis. The reasons in favour of proposing inductive underlying processes are drawn from the ill-defined problems in our environment, the need to manage a wealth of information and, finally, the value to produce effective and efficient conclusions (and solutions) to everyday problems. The inductive hypothesis is explored in this paper by (a) training an inductive architecture such as a connectionist architecture to solve a traditional formal task (e.g., Wason's selection task), (b) manipulating specific inductive variables in both an abstract and a thematic version of Wason's task, and (c) conducting think aloud interviews with participants as they solve a thematic version of Wason's task. Results suggest that a connectionist architecture is able to generate a solution to Wason's task, and its solution offers an alternate perspective to formal theories of performance on this task. Furthermore, results suggest that manipulating inductive variables *does* influence performance on an abstract version of Wason's task, but does *not* influence performance on a thematic version of the task. Finally, responses obtained from the think aloud interviews suggest that participants adopt inductive strategies to solve a thematic version of Wason's task. Future "inductive" studies of thematic effects may be unable to disentangle familiarity or background knowledge from basic inductive reasoning processes. The implication of these results are discussed.

Acknowledgment

This thesis would not have been possible without the assistance of several people. I would like to thank my co-supervisors, Professor Michael R.W. Dawson and Professor C. Donald Heth for guiding me through new intellectual frontiers and asking me tough questions along the way. I would especially like to thank Professor Dawson for providing a challenging lab atmosphere where ideas were continually offered and debated, and where traditional methods were questioned as rigorously as innovative ones. I learned the most during these conversational exchanges. I would also like to thank Professor Gay Bisanz for scrutinizing the ideas held in the thesis and offering insightful suggestions wherever needed. I also want to acknowledge the Social Sciences and Humanities Research Council of Canada (S.S.H.R.C.) for funding this research. Finally, I would like to thank my husband Greg who helped me navigate through the many impasses research presents and who believes as much as I do in the purpose of research. Without belief in a greater purpose, then what we do is just done in vain.

Table of Contents

List of Tables

List of Figures

Chapter 1

An Alternate Approach to Understanding Formal Reasoning: Thinking According to the Inductive-Coherence Model

Introduction

In the past 40 years, psychologists have debated both the rationality and the *underlying processes* of human reasoning. In this debate, the underlying processes of reasoning are differentiated from reasoning *strategies*. Underlying processes is a term used to describe the "functional architecture" of reasoning (Dawson, 1998). The identification of the functional architecture is important because it anchors functional theories of cognition to more concrete descriptions of cognitive phenomena (Dawson, 1998). In this respect, an identification of the functional architecture of human cognition provides "an account of the mental programming language in which cognitive algorithms are written" (Dawson, 1998). In contrast, reasoning strategies emerge from the functional architecture. One way to differentiate underlying processes or functional architecture from strategies is to think of underlying processes as being removed from consciousness awareness, and of strategies as being accessible to conscious awareness.

Although the debate is generally about human reasoning, it centres on *formal reasoning* because this is the kind of reasoning that has been the focus of most psychological research (Galotti, 1989; Garnham & Oakhill, 1994). Formal reasoning may be defined as reasoning that is used in formal domains such as in mathematics or computer programming; a logical, and highly systematized method for drawing inferences (Garnham & Oakhill, 1994). Many psychologists believe that people use formal reasoning when drawing inferences in response to mathematical problems or logical tasks (e.g., Braine, 1978; Braine & O'Brien, 1991; Galotti, Baron, & Sabini, 1986; Rips, 1994). In general, the label "formal reasoning" is used to describe deductive inference. Moreover, formal reasoning is typically set apart from the everyday reasoning people employ to solve less structured problems such as deciding which route to take home on a busy Friday

afternoon. Although investigators have neglected to study everyday reasoning empirically, they believe that everyday reasoning may be predicted from performance on formal reasoning tasks (Galotti, 1989; Garnham & Oakhill, 1994).

Prior to the debate, psychologists generally believed that human reasoning followed formal, logical rules (Henle, 1962). However, this belief began to deteriorate when investigators started to test human participants on logical tasks such as *categorical syllogisms* and *conditional syllogisms* (Henle, 1962; Janis & Frick, 1943; Woodworth & Sells, 1935). While it is difficult to say precisely when the debate over both the rationality and the character of the underlying processes of human reasoning began, the debate was ignited by Peter Wason and the experimental results he obtained from his card selection task (1966). Peter Wason devised his card selection task as a test of people's formal reasoning capabilities; in particular, people's hypothesis-testing capabilities. *Wason's selection task* is described fully in Chapter two, but suffice it to say for now that when Wason administered his task to college students, most performed very poorly. Specifically, over 90 percent of students failed to test hypotheses according to accepted formal methods of hypothesis-testing. This result challenged the belief of human rationality and the related belief that human reasoning emerged from formal underlying reasoning processes. Wason's task is not the only formal task that has evoked poor performance from human participants. Other tasks such as categorical syllogisms and conditional syllogisms have also been used by investigators to show that human reasoning is faulty (e.g., Lefford, 1946).

A number of investigators have proposed theories to explain participants' poor reasoning on formal tasks (e.g., Braine, 1978; Braine & O'Brien, 1991; Cheng & Holyoak, 1985; Cosmides, 1989; Evans, 1989; Johnson-Laird & Byrne, 1991; Rips, 1994). For example, Rips (1994, 1995) maintains that people possess formal underlying reasoning processes, which give rise to deductive strategies. He accounts for people's poor formal reasoning by claiming that specific endogenous factors *interfere* with the operation of formal underlying processes. The two endogenous factors include *ordinary comprehension processes* and *working memory limitations*. Although there is nothing

ordinary about people's comprehension processes, the term "ordinary comprehension processes" is used in the literature to identify people's everyday, pragmatic interpretation of information; interpretations that often lead to invited inferences, which are not logically necessary (Braine, 1978; Braine & Rumain, 1983; Garnham & Oakhill, 1994). Ordinary comprehension processes are distinguished from "analytic comprehension processes," which lead to necessary inferences, and are not typically used in everyday settings.

In contrast, other theorists maintain that people possess *informal* underlying reasoning processes. The label "informal underlying reasoning processes" is a general name used to describe inductive processes (Galotti, 1989; Garnham & Oakhill, 1994). The purpose of inductive processes follows Rescher's (1980) interpretation:

> Induction is at bottom a mechanism for enlarging the stock of (purported) truths that we accept about the world -- a resource comparable in this regard to observation and memory .... Induction is an instrument for question-resolution in the face of imperfect information. It is a tool for use by finite intelligences, capable of yielding not the best *possible* answer (in some rarified sense of this term), but the best *available* answer, the best we can manage to secure in the existing conditions in which we do and must conduct our epistemic labors. (p.6-7)

Because induction, according to Rescher (1980), is used to reason within everyday domains, inductive reasoning may also be termed *everyday* reasoning.

Many theorists who advance an informal account of human reasoning *also* claim that participants' poor formal reasoning is due to interference from ordinary comprehension processes and working memory limitations (e.g., Cheng & Holyoak, 1986; Johnson-Laird & Byrne, 1991). For example, Johnson-Laird and Byrne (1991) claim that people reason poorly when they misinterpret the information presented in the problem or fail to explore multiple interpretations; thereby leading to inadequate problem solving strategies. Similarly, Cheng and Holyoak (1985) claim that adequate problem solving requires participants to interpret the problem information properly in order for the appropriate schema to be invoked. However, the assumption that ordinary comprehension processes and working memory limitations interfere with inductive underlying reasoning processes is

problematic because both these factors should facilitate inductive reasoning processes (Holland, Holyoak, Nisbett, & Thagard, 1986). For instance, Holland et al., (1986) suggest that ordinary comprehension processes and working memory limitations constrain how inductive reasoning processes search the problem space, thereby facilitating problem solving. As a result, theorists who advance an informal account but who also suggest that poor reasoning on formal tasks arises from interfering factors fail to offer a theoretically different account of formal reasoning from those theorists who suggest people possess formal underlying processes.

Although existing evidence can be interpreted to support the hypothesis that people possess formal underlying processes that go awry due to interfering factors (i.e., ordinary comprehension processes and working memory limitations), this same evidence can also be interpreted to support an alternative hypothesis: People make mistakes on formal tasks because they possess inductive underlying reasoning processes that are incompatible with standard formal tasks. The incompatibility comes from the fundamental partnership between inductive processes and both ordinary comprehension processes and working memory limitations. Ordinary comprehension processes and working memory limitations function to narrow the problem space in order for inductive processes to generate efficient strategies for problem solving (Holland et al., 1986). However, ordinary comprehension processes and working memory limitations are not interfering factors for informal processes; they are fundamentally necessary to the operation of inductive processes (Holland et al., 1986). Ultimately, judging the adequacy of either the formal or informal hypothesis will rest on empirical evidence but also to some extent on arguments of evolutionary adaptation. While evolutionary arguments are tricky to make since they are necessarily post hoc, they are important to at least consider since human behaviour is essentially tied to its environment.

The central theme of this paper is inspired by words from Newell's and Simon's (1972) influential analysis of human problem-solving: "[T]o the extent that the behavior departs from perfect rationality, we gain information about the psychology from the subject, [and] about the nature of the internal mechanisms limiting his performance"

(p.55). The central theme of this paper is to present *an explicitly informal account of participants' performance on traditional formal reasoning tasks.* This presentation requires examining participants' formal performance within an everyday problem solving framework. Within this framework, participants' underlying reasoning processes are assumed to be inductive. Underlying inductive processes may give rise to either inductive or deductive reasoning strategies, but deductive strategies will arise primarily through formal training and only in response to highly specific domains such as mathematics. As such, inductive strategies are expected to characterize unrehearsed reasoning–the kind of reasoning measured by standard formal tasks used in psychological experiments. Consequently, biases and errors will result when people attempt to solve formal tasks with "inductive" strategies.

An explicitly informal account of participants' performance on formal tasks has heretofore not been offered in the literature. As the reader will recall, even informal accounts implicitly point to the existence of formal underlying reasoning processes (e.g., Johnson-Laird & Byrne, 1991). The hypothesis advanced here is new in that inductive underlying reasoning processes instead of formal underlying reasoning processes are explicitly proposed. The reasons in favour of proposing inductive underlying reasoning processes stem from (a) the ill fit between the assumption of formal underlying reasoning processes and participants' performance on formal tasks, and (b) the ill-defined quality of the problems in our everyday environment and the adaptive character (i.e., necessity to manage a wealth of information and generate efficient and effective solutions) of human information processes, both of which point to the necessity of inductive processes.

A Road Map to the Present Paper

To fulfil the goal of presenting an informal account of participants' performance on traditional formal reasoning tasks, the present paper is organized in three main sections. In the first section, I review the research studies that have led to the hypothesis that individuals possess formal reasoning processes, and the associated hypothesis that interfering factors are responsible for the errors observed on formal tasks. Most, if not all,

research studies that are reviewed involve one of three standard formal tasks–categorical syllogisms, conditional syllogisms, and Wason's selection card task (Wason, 1966). The intent of this section is to (a) demonstrate that most of the theories offered to explain formal reasoning errors, explicitly or implicitly, endorse the existence of formal underlying reasoning processes; and (b) offer a new informal account of reasoning based on an *inductive-coherence* model of reasoning.

Given a rationale and model for an informal account of reasoning, in the second section I demonstrate that an inductive architecture such as a connectionist architecture can solve Wason's selection task, and how such a solution is organized. The intent of this section is to show that the appearance of "formal" performance (e.g., solution of Wason's task) need not rely on a formal architecture but can emerge from an informal architecture.

After demonstrating that formal performance can originate from informal processes, in the third section, I manipulate *specific* inductive variables within a version of Wason's task in order to clearly show that formal performance is controlled via inductive processes. Although another inductive variable, namely content, has been manipulated in the past and shown to alter participants' performance on Wason's task, the problem with this manipulation is that it supports not only an informal account of reasoning but also a *memory or familiarity* account of reasoning. Proponents of this latter account suggest that content can facilitate or hinder formal performance depending on the memories the content cues in participants. My intent in this section is to rule out the memory account of performance by manipulating variables that do not influence memory recall but instead influence only inductive processes. Finally, future avenues of research are suggested for testing inductive hypotheses of formal reasoning performance.

Chapter 2

Inductive Underlying Reasoning Processes and Performance on Formal Tasks

Reasoning, the ability to form conclusions, judgements, or inferences based upon evidence, is central to human cognition. Reasoning characterizes the high-level mental activity that humans must employ if they are to function successfully within a complex environment (Garnham & Oakhill, 1994). People constantly reason about objects, events, and other people in order to decide which beliefs and actions will bring them closer to their goals.

Goals range in importance. Some goals are fairly mundane such as deciding which route leads to the quickest passage home. Other goals are noteworthy in that they involve high stake decisions. For instance, imagine the extraordinary task jury members face in deciding whether a person is guilty of a crime based on the available evidence. What evidence, from the multitude of possibilities, is considered relevant? How is the evidence weighed? How does the evidence map onto the judgement or conclusion? Sound reasoning is essential in high-stake tasks, but it is also essential in ordinary tasks where individuals have unique goals to attain.

A multitude of studies, however, has led researchers to conclude that human reasoning does not always lead to logical conclusions (e.g., Begg & Harris, 1982; Evans, 1989; Evans, Newstead, & Byrne, 1993; Griggs, 1983; Janis & Frick, 1943; Johnson-Laird & Byrne, 1991; Markovits, 1985; Wason, 1966; and Wason, 1983). In particular, people typically respond incorrectly to *formal* (or *deductive*; the two terms are used interchangeably here) tasks such as categorical and conditional syllogisms (e.g., Revlin & Leirer, 1978; Rumain, Connell, & Braine, 1983; and O'Brien, 1993). Although formal tasks are dissimilar to the *informal* (or *everyday*) tasks that people normally encounter, performance on formal tasks is believed to predict performance on informal tasks (Galotti, 1989). It is assumed that the reasoning processes elicited by formal tasks are similar, if not

the same, to the processes elicited by everyday tasks. Hence, a participant's poor performance on a formal task is taken as an indicator of his or her poor performance on an everyday task. Despite the equivalence assumed of formal and everyday tasks, the few studies that have explored the relationship between formal and everyday task performance suggest, at best, a weak correspondence (see Galotti, 1989). Far more study needs to be devoted to the relation, if any, between formal and informal reasoning processes.

Reasoning errors on formal tasks have shed doubt on the rationality of human reasoning. While few scholars today believe that people possess faulty reasoning processes (Cohen, 1981; Harman, 1986, 1995; and Johnson-Laird & Byrne, 1991), many hypothesize that human reasoning can, and frequently does, go awry and yield faulty conclusions to logical tasks. The discrepancy between sound reasoning processes on the one hand, and faulty performance on the other hand is reconciled by postulating the existence of *interfering factors* that disrupt reasoning processes (see Evans, 1989). Interfering factors include (a) *ordinary comprehension processes* or the use of pragmatic instead of analytical comprehension processes to interpret language (Grice, 1975); and to a lesser extent (b) *working memory limitations* (see Braine, 1978; Henle, 1962; Hitch & Baddeley, 1976; Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1991; and Markovits & Guilaine, 1989). Although the literature does provide some evidence that interfering factors can disrupt reasoning, another hypothesis worth considering is that errors observed on formal tasks result from informal or inductive underlying reasoning processes, which are applied to formal tasks. Indeed, as noted previously, the goal of this paper is to present an explicitly informal account of participants' performance on traditional formal reasoning tasks. This presentation requires examining participants' formal performance within an everyday problem solving framework. Within this framework, participants' underlying reasoning processes are assumed to be inductive and, as such, are mismatched with the design of formal tasks. Consequently, biases and errors result when participants attempt to solve formal tasks with "inductive" strategies.

Interpreting formal performance by way of inductive underlying processes is not a new idea. A number of investigators have either hinted at or fully advanced this idea (e.g.,

Barsalou & Goldstone, 1998; Cosmides & Tooby, 1996; Evan et al., 1993; Liberman & Klar, 1996; Oaksford & Chater, 1994; Staudenmayer, 1975; Sloman & Rips, 1998; Stevenson & Over, 1995). Indeed, some investigators have developed influential reasoning theories based *implicitly* on the inductive treatment of formal tasks. At the core of these theories, however, *formal* reasoning errors are *still* explained in traditional formal terms, viewing them as deviations arising from the existence of interfering factors (e.g., Cheng & Holyoak, 1985; Johnson-Laird & Byrne, 1991). Consequently, these alternative theories represent neither distinct alternatives to the "mental logic" approach nor do they fully advance an inductive approach to reasoning performance.

As mentioned already, the interfering factor hypothesis of people's poor performance has received some empirical support (e.g., Gilhooly, Logie, Wetherick, & Wynn, 1993; Rumain, et al., 1983), but this support is not definitive. Assuming the interfering factor hypothesis, how is it possible that people are able to effectively meet goals and draw conclusions in everyday tasks, which one would assume are also subject to interfering factors (Galotti, 1989)? Indeed, as Rescher (1980) comments, human reasoning has so far survived evolutionary pressures so our reasoning system cannot be all that incompetent. One possible answer to this question is that interfering factors do not disrupt reasoning because formal processes do not characterize our underlying reasoning processes. Perhaps, the underlying processes of human reasoning are better viewed as informal rather than formal.

The reasons that human reasoning processes are more likely to be informal instead of formal stem from the nature of the environment. First, the problems encountered in everyday settings are typically *ill-defined*, meaning they lack a complete and clear specification of the information that is required to solve them (Dawson, 1998; Galotti, 1989; Garnham & Oakhill, 1994; Ginsberg, 1986; Holland et al., 1986). Individuals presumably solve these ill-defined problems by "filling in" the missing information with background or *default* information; that is, information extracted from the stable, and unchanging properties of the environment (Rescher, 1980). Second, assuming that problems in the environment are ill-defined, the conclusions or inferences drawn to solve

such problems need to be defeasible or, in other words, flexible in light of further information (Collins & Michalski, 1989). Formal processes, however, cannot produce defeasible conclusions because they, by definition, need to operate on complete information, which ultimately leads to definitive conclusions (Ginsberg, 1986). Third, the reasoning processes most likely to operate successfully in an environment of ill-defined problems must be constrained or limited so that solutions can be found efficiently (Fodor, 1983; Holland et al., 1986). For instance, such constraints might include (a) capitalizing on a person's background knowledge (i.e., the use of ordinary comprehension processes help one to focus on information selectively, thereby helping to reduce the problem space), including his or her goals, (b) producing plausible conclusions instead of definitive conclusions, and (c) minimizing the problem space (i.e., consider single solutions instead of many alternative solutions) (Holland et al., 1986).

If individuals are to function effectively in an ill-defined environment, they must increase information in the presence of the uncertainty associated with ill-defined tasks (Bisanz, Bisanz, & Korpan, 1996; Holland et al., 1986). Inductive underlying processes can accomplish this goal because they exploit a reasoner's prior knowledge (i.e., inductive processes rely on ordinary comprehension processes to focus on specific information), draw defeasible conclusions, and narrow the problem space. To be sure, one drawback associated with inductive processes is that their application to formal tasks will result in reasoning errors. Errors will occur because the requirements associated with formal tasks conflict with the constraints of inductive processes. For example, formal tasks require participants to neglect prior beliefs, generate definite conclusions, and explore all possible solutions before choosing the best solution. In contrast, inductive processes exploit prior knowledge, generate defeasible conclusions, and consider a narrow problem space. Errors on formal tasks then might occur because individuals apply inductive processes to non-inductive tasks. This inductive account of formal errors is a different explanation of formal errors than one which assumes that interfering factors disrupt formal processes.

Overview of Chapter Two

In this chapter, I first present an *inductive-coherence* model for understanding reasoning. Second, I present a review of the research studies conducted with three standard formal reasoning tasks: Categorical syllogisms, conditional syllogisms, and Wason's selection task. As each task is reviewed, the formal process account of performance on the task is first provided, followed by the inductive account. Finally, the inductive process account is revisited with concluding remarks.

Everyday Tasks, Increased Information, and Induction: An Inductive-Coherence Model of Reasoning

In contrast to deduction, induction does not permit the individual to draw absolute conclusions from a set of true premises. Conclusions arrived at by means of induction are equivocal or "non-truth preserving" because they extend beyond the information provided in the premises and so risk being incorrect. Although such conclusions are indefinite, they increase the reasoner's semantic information. In contrast, truth-preserving conclusions arrived at by means of deduction do not increase semantic information. Paradoxically, it is the very indefiniteness of inductive conclusions that permits us to solve everyday problems because such conclusions facilitate our ability to increase the information obtained from the premises presented or experience (Rescher, 1980).

The ability to form conclusions that increase our semantic information is important in an environment of ill-defined problems and extensive problem-spaces (Bisanz et al., 1994; Dawson, 1998; Holland et al., 1986; Thagard, 1986; Neisser, 1982). Many environmental tasks are ill-defined (Dawson, 1998; Galotti, 1989; Neisser, 1982) because they lack the information required to solve them; that is, they cannot be solved simply by attending to the information presented in the task. In order to solve the task, one must supplement the problem's information with other, additional details. In addition, because ill-defined tasks may be solved in principle using an infinite number of approaches, the reasoner needs to reduce this problem space by invoking additional information.

As Holland et al., (1986) indicate, inductive reasoning facilitates solving ill-defined

tasks by (a) exploiting the reasoner's background information, (b) drawing non-monotonic conclusions (i.e., tentative conclusions that can be reversed or retracted in light of new environmental information), and (c) using these initial conclusions as starting points to problem solving, thereby reducing the problem-space. Using these three techniques, inductive reasoning allows reasoners to generate appropriate solutions swiftly (Ginsberg, 1986; Harman, 1986, 1995).

The everyday environment of reasoners needs to be incorporated into a theory of human reasoning since the products of reasoning are ultimately applied to solve environmental problems. For this reason, theorists need to consider how particular reasoning processes might be adapted to the reasoners' world (Cosmides & Tooby, 1996). Hence, underlying reasoning processes (i.e., reasoning architecture) premised on deductive or logical rules seem unlikely if our environment contains, and reasoners need to solve, ill-defined problems. Formal processes neglect those very features that inductive processes exploit to allow reasoners to solve ill-defined problems. Formal processes do not exploit background information and by definition do not draw non-monotonic conclusions.

To be sure, reasoners do not just solve ill-defined problems. Reasoners also solve deductive problems (e.g., mathematical problems) but these do not represent the bulk of the everyday problems for which our reasoning system likely adapted (Rescher, 1980). A system of inductive processes is the most plausible candidate for representing human reasoning because, as Holland et al., (1986) note, "[it] is subject to constraints that can be derived from the general nature of an information-processing system that pursues goals in a complex environment and receives feedback about its successes in attaining its goal (p.5).

Using coherence theory (Gärdenfors, 1991) and Holland et al's (1986) description of the features associated with inductive processes, I propose an *inductive-coherence* model of reasoning for understanding the errors observed on formal tasks. As shown in Figure 2.1, the top of the model illustrates the reasoner's environment, including its everyday problems along with the appearance of additional information with the passage of time.

The circle in the middle of Figure 2.1 represents the reasoner. The various shapes inside the circle reflect the reasoner's distributed knowledge or beliefs (including goals). In general, in the inductive-coherence model of reasoning, it is assumed that problems in the environment are ill-defined and so, to be solved, they must evoke a reasoner's background knowledge. At the same time, the reasoner imposes his or her own background knowledge on the problem to guide and promote efficient problem solving (i.e., reduce the problem space).

If the problem lacks meaningful content and, therefore, fails to evoke background knowledge and/or the reasoner cannot impose this knowledge, then reasoning will break down. In this case, reasoners will use simple heuristics to generate a response. For example, participants have been observed to apply simple heuristics such as matching the surface features of the conclusion to the surface features of the premises in response to abstract formal tasks (e.g., Begg & Denny, 1969; Evans & Lynch, 1973). In contrast, problems rich in meaningful content can lead reasoners down one of two roads. On the one hand, problems rich in meaning may evoke very specific knowledge or *single solutions* if there are not any compelling reasons to consider alternative solutions. Investigators have observed repeatedly that participants do not consider alternative solutions to a problem when a proposed solution supports participants' existing attitudes or beliefs about the world (Janis & Frick, 1943; Newstead, Pollard, & Evans, 1992; Sloman, 1994; Swain, 1979). This problem solving route, which involves single solutions, is called the *path of least resistance* and is the path that explains why most reasoners perform poorly on formal tasks that are rich in meaning or "thematic." Essentially, many thematic formal tasks either do not challenge participants' view of the world or represent trivial problems. As a result, these tasks evoke the single most available piece of knowledge from a reasoner's memory, prompting the reasoner to respond quickly and according to this information. In the absence of compelling reasons to consider alternate information, participants accept the single solution; that is, they reason according to the path of least resistance because they lack compelling reasons to consider alternative solutions. In many ways, reasoning according to the path of least resistance represents an

economical approach to information processing because it does not expend energy analysing conclusions that are known to be true or are too trivial to examine closely.

On the other hand, problems that challenge participants' view of the world or that prompt participants to attend to information that contradicts the information presented in the problem are solved by a secondary path. Because different alternative pieces of information are evoked in this secondary path, the reasoner is likely to doubt or experience a decreased sense of confidence about a quick and easy solution to the problem. This decreased sense of confidence promotes in-depth but inefficient problem solving because the reasoner has to sift through the available information and reconcile any inconsistencies. It is in the secondary path where logical performance originates, not from deductive rules but from the weighing of alternatives. A solution generated either from the path of least resistance or the secondary path can be retracted and modified if additional information is found to negate the conclusion. This model is used to understand the reasoning errors found on categorical and conditional syllogisms, as well as Wason's selection task.

Categorical and Conditional Syllogisms, and Wason's Card Selection Task: Formal Versus Informal Reasoning Tasks

Most researchers who study reasoning on formal tasks do so because they ultimately expect it to predict reasoning on everyday tasks (Galotti, 1989). As mentioned previously, however, no strong relationship between performance on formal tasks and everyday tasks has been established (Galotti, 1989). Nonetheless, investigators continue to study formal reasoning tasks (instead of everyday tasks) because these tasks are linked to an established research methodology (Garnham & Oakhill, 1994). For instance, in formal tasks the investigator can control the information participants use to generate a solution because participants do not need information outside of that provided in the task. In contrast, in everyday, ill-defined tasks, investigators cannot exercise such control because part of the information needed to solve the task exists outside of the task (Galotti, 1989; Garnham & Oakhill, 1994). As a result, formal tasks have taken center stage in much of reasoning

research, along with the assumption that people possess formal underlying reasoning processes. The three standard formal tasks that have dominated reasoning research include *categorical syllogisms*, *conditional syllogisms*, and *Wason's selection task* (Wason, 1966).

Theories of human reasoning have been developed in part as a result of participants' performance on categorical syllogisms, conditional syllogisms, and Wason's selection task (e.g., Cheng's & Holyoak's (1986) pragmatic reasoning theory and Johnson-Laird's & Byrne's (1991) mental models theory). Rips' (1994) theory, however, did not grow out so much from results obtained on standard formal tasks as much as from working assumptions about the nature of our reasoning architecture. Proponents of these theories hold interfering factors (i.e., ordinary comprehension processes and working memory limitations) as responsible for the errors observed on formal tasks. Blaming interfering factors for these errors, however, presumes that without them formal reasoning processes should operate and yield logical conclusions. It is possible, however, that people commit errors on formal tasks because they possess inductive reasoning processes that have specific constraints geared to deal with the ill-defined problems present in our everyday environment. If the errors committed on formal tasks are the products of inductive constraints, they should not be viewed as errors but as reasonable responses arising from inductive processes.

Categorical Syllogisms

Task Definition

One of the three standard tasks used to study formal reasoning includes categorical syllogisms. Categorical syllogisms consist of two quantified premises and a quantified conclusion. The premises reflect an implicit relation between a *subject* (S) and a *predicate* (P) via a *middle term* (M), while the conclusion reflects the relation between S and P explicitly. Both the premises and the conclusion are normally structured in one of four *moods*. The moods are identified with the letters *A* (i.e., All A are B), *I* (i.e., Some A are B), *E* (i.e., No A are B), and *O* (i.e., Some A are not B). Moods A and E reflect universal

relations, whereas moods I and O reflect particular relations.

Another feature of the categorical syllogism is that its premises and conclusion must be organized in one of four *figures*. A figure characterizes the location of the subject, predicate, and middle term. Figure 2.2 shows the four traditional figures of a syllogism. Given that a categorical syllogism must be in one of four figures and that its premises and conclusion must each be in one of four moods, 256 different syllogisms can be created (i.e., 4 x 4 x 4 x 4 = 256). Of these 256 syllogisms, only 24 are considered valid. The validity of these 24 syllogisms may be proven using either proof-theoretic methods or, more commonly, a model-theoretic method, in which a valid syllogism is one whose premises cannot be true without its conclusion also being true (Garnham & Oakhill, 1994). This latter method of proving validity requires that all possible interpretations of the premises be considered along with the putative conclusion. If the conclusion holds in all interpretation of the premises, then the syllogism is valid. Figure 2.3 shows the 24 valid syllogisms from the group of 256 syllogisms.

Reasoning studies employing categorical syllogisms normally use either abstract or thematic syllogisms. Abstract syllogisms typically use letters to represent the subject, predicate, and middle terms (e.g., All A are B), whereas thematic syllogisms use nouns or other meaningful terms (e.g., All cooks are painters). Participants are normally required to either generate an appropriate conclusion to a set of premises, judge whether a presented conclusion follows from a set of premises, or choose the conclusion that best follows from a set of premises from a list of alternatives.

Biases Observed on Abstract Categorical Syllogisms

People normally commit errors on abstract categorical syllogisms. For example, people typically endorse or draw conclusions that match the "atmosphere" (i.e., draw conclusions that are similar to the surface features of the premises) of the syllogism's premises, irrespective of whether the endorsed conclusion logically follows (Begg & Denny, 1969; Woodworth & Sells, 1935). Such errors have led to a number of explanations about the processes that people engage when they attempt to solve categorical syllogisms. Some of

these explanations include (a) the *atmosphere hypothesis* (Woodworth & Sells, 1935) and (b) *availability theory* (e.g., Pollard, 1982), both of which are strategic or heuristic accounts, and (c) *conversion theory* (Revlin & Leirer, 1978), (d) *mental models theory* (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991), and (e) *syntactic or "mental logic" theory* (e.g., Rips, 1994), all of which are process accounts. These five explanations are discussed in more detail below.

According to the atmosphere hypothesis, when at least one of the premises is negative (i.e., a premise in either moods E or O), participants tend to accept a negative conclusion from a list of alternative conclusions. Likewise, when at least one of the premises is particular (i.e., a premise in moods I or O), participants tend to accept a particular conclusion from a list of alternatives. Early investigators (for a review see Evans, Newstead, & Byrne, 1993) viewed this biased performance as *prima facie* evidence for the irrationality of human reasoning processes.

Endorsing a conclusion that is similar to the surface features of its premises has also been attributed to an availability heuristic. Drawing from the research in statistical reasoning pioneered by Tversky and Kahneman (1974, 1982), Pollard (1982) holds that individuals accept conclusions that are similar to their premises because such conclusions are easily and quickly conceived. The atmosphere hypothesis and availability theory are explanations that essentially hold interfering factors such as ordinary comprehension processes responsible for skewing task performance because reasoning errors arise from participants' biased interpretation of task information. Proponents of these accounts do not explain how underlying reasoning processes should in principle operate. To be sure, these accounts do a reasonable job predicting responses (Begg & Denny, 1969; Newstead, 1989). Their main weakness is that they fail to provide an account of why participants sometimes endorse the "no valid conclusion" alternative (Ceraso & Provitera, 1971; Evans et al., 1993; Garnham & Oakhill, 1994). These accounts are also weakened by studies that suggest participants respond in accord with the atmosphere of the premises only when this response is also logically correct (Ceraso & Provitera, 1971). The recognition that people are able to endorse or draw logically correct conclusions in response to formal tasks under

some conditions has led many investigators to reject the idea that human reasoning processes are inherently faulty (e.g., Ceraso & Provitera, 1971; Henle, 1962). Instead, many embrace the idea that people can reason logically if they are able to properly interpret the premises and avoid interfering factors such ordinary comprehension processes and working memory limitations (Begg & Harris, 1982; Ceraso & Provitera, 1971; Henle, 1962; Revlin & Leirer, 1978; and Revlin, Leirer, Yopp, & Yopp, 1980).

Proponents of conversion theory (Revlin & Leirer, 1978; Revlin et al., 1980), for instance, hold that illogical responses are not due to faulty reasoning processes but instead due to people's idiosyncratic encoding of the premises. Conversion theorists claim that during encoding participants normally convert one or both premises of a syllogism such that, for example, "All A are B" is interpreted to mean "All B are A." Except for a few categorical syllogisms, such conversions will lead to logically incorrect conclusions. Hence, illogical responses, according to conversion theorists, are only illogical according to the investigator's interpretation of the premises and not according to the participant's interpretation of the premises. Attributing reasoning errors to people's encoding of premises, however, basically attributes errors to ordinary comprehension processes. By holding ordinary comprehension processes as responsible for hindering logical reasoning, it is implicitly assumed that, without them, logical reasoning would proceed unencumbered.

Although conversion theorists are much less influential today, the comprehension issue they struggled with is strangely similar to the one influential theories today must reconcile, namely, how comprehension of task information distorts reasoning performance. Johnson-Laird (1983; Johnson-Laird & Byrne, 1991) is a good example of an influential theorist wrestling with this issue today. In his mental models theory, individuals construct a model of the premises and, within the model, attempt to generate a possible conclusion. Although the construction process is not fully explained, he maintains that reasoning errors occur because participants do not exhaustively explore all potential models or *interpretations* of the premises. Consequently, the conclusion derived under one model may be indefensible given another, not considered by the participant but equally plausible given the premises.

It is proposed that one reason that participants do not search exhaustively for alternate models is due to either ordinary comprehension of the task information or to limited working memory capacity or both (Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1991). For example, if participants construct a single model of the premises "Some A are B" and "No B are C," and ignore other possible interpretations, then it is highly likely that they will generate a wrong conclusion because they have failed to consider if their conclusion holds in the other models capable of being constructed for the premises. Working memory can also disrupt the search for alternate models if the number of models that needs to be considered exceeds working memory limitations. In support of the interfering factor hypothesis, studies of working memory and categorical reasoning suggest that the central processor or executive and, to a lesser degree, the articulatory loop are involved in syllogistic reasoning (e.g., Gilhooly, Logie, Wetherick, and Wynn, 1993). For instance, Gilhooly et al., (1993) present evidence to show that the accuracy and processing time for solving a categorical syllogism is reduced when participants simultaneously engage in solving the syllogism task and a secondary task designed to load the central executive.

Finally, syntactic theorists, who claim that people reason according to rules that operate on the syntax of the premises, argue that individuals perform poorly on abstract categorical syllogisms because they lack either the appropriate logical rules, the analytical comprehension processes, or the working memory capacity to deduce the valid response (e.g., Braine & Rumain, 1983; Rips, 1994). Implicating ordinary comprehension processes and working memory limitations as reasons for logical reasoning errors assumes that people possess a formal reasoning architecture since these "interfering factors" are hypothesized to facilitate an informal reasoning architecture (Holland et al., 1986).

Biases Observed on Thematic Categorical Syllogisms

People do not only commit errors on abstract categorical syllogisms; people also commit errors on thematic categorical syllogisms, whose meaningful content one would expect to aid performance. Errors on thematic categorical syllogisms, which are normally

referred to as *belief bias* errors, have also yielded a number of interpretations about the processes people employ when they solve such tasks. Belief bias is the tendency to reject unbelievable conclusions and to accept believable conclusions, irrespective of whether they logically follow from the premises (Janis & Frick, 1943; Lefford, 1946). Participants who exhibit belief bias appear to ignore the instructions of the task, and instead to base their responses on background knowledge and belief rather than logic. Using prior knowledge or belief to solve categorical syllogisms is considered undesirable because it can lead to errors. In principle, formal tasks must be solved using only the information provided in the premises because background knowledge is believed to interfere with the inferential process by either altering the way in which the premises are interpreted or by making logically unnecessary conclusions appear necessary.

In addition to a general effect of belief bias, investigators have observed a *belief by logic* interaction. This interaction reveals that participants make more errors on invalid syllogisms than on valid syllogisms when the conclusion is believable, but few errors on both invalid syllogisms and valid syllogisms when the conclusion is unbelievable (Evans, Barston, & Pollard, 1983; Evans, 1989). Many theorists hold ordinary comprehension processes responsible for these effects (e.g., Evans et al., 1983; Evans, 1989; Johnson-Laird & Byrne, 1991; and Revlin & Leirer, 1978). Some of the theories include conversion theory (Revlin & Leirer, 1978), the *selective scrutiny model* (Evans, 1989), the *misinterpreted necessity model* (Evans, 1989; Newstead, Pollard, Evans, & Allen, 1992), and mental models theory (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991).

For example, Revlin and Leirer (1978), in their conversion theory, claim that meaningful content influences participants' interpretation of the premises, including whether they convert the premises. Content that induces participants to convert the premises of an argument is associated with yielding a greater proportion of incorrect conclusions. In contrast, content that discourages participants from converting premises is associated with yielding a greater proportion of correct conclusions. Conversion of premises, however, has been found to play a minor role in the explanation of belief bias (Newstead, 1989). For example, when syllogisms whose conclusions remain unaffected by

conversion of premises are used as tasks, individuals continue to endorse believable conclusions. This result suggests that there is something else causing participants to endorse believable conclusions than just conversion of premises. Furthermore, conversion theorists have not satisfactorily accounted for the interaction between belief and logic (Newstead, Pollard, Evans, & Allen, 1992).

In an effort to account for the interaction between belief and logic, Evans (1989) and Newstead et al., (1992) proposed the *selective scrutiny model* and the *misinterpreted necessity model*, respectively. Proponents of the selective scrutiny model account for belief bias from a heuristic perspective because they suggest that individuals inspect the logic of the syllogism only when its conclusion is unbelievable. Participants therefore should make few mistakes on both valid and invalid unbelievable syllogisms because only under these circumstances do they expend the effort to inspect the logic of the syllogism. In contrast, participants should make many errors on invalid syllogisms when the conclusion is believable because they would accept the conclusion without inspecting the syllogism's logic. Also, participants should make few errors on believable valid syllogisms because accepting these conclusions at face value results in correct responses.

Proponents of the misinterpreted necessity model account for belief bias effects from a process perspective because they posit that individuals work through an argument in order to formulate a conclusion that is either determinately true or false. According to this model, participants should commit few errors on valid syllogisms because true determinate conclusions can be derived; however, participants should commit many errors on invalid syllogisms because indeterminate conclusions are frequent. When participants attempt to generate a determinately false conclusion to an invalid syllogism and find they are unable (because indeterminate conclusions are numerous), they will fall back on a belief heuristic in which believable conclusions are accepted and unbelievable conclusions are rejected at face value. Ordinary comprehension processes are thus implicated under the misinterpreted necessity model since whether subjects fall back on a belief heuristic depends on their interpretation of the premises, and their subsequent ability to generate a conclusion.

After conducting experiments to decide which model was better able to account for the belief by logic interaction, Newstead et al., (1992) found that only the misinterpreted necessity model was able to explain why participants generated fewer biased responses to invalid categorical syllogisms associated with determinately false instead of indeterminately false conclusions. According to the misinterpreted necessity model, a determinately false conclusion is easier to reject because participants can match their own self-generated "false" conclusion against the syllogism's presented false conclusion. In contrast, an indeterminately false conclusion is more difficult for participants to reject because these conclusions do not always match participants' self-generated false conclusions, leading participants to commit more errors.

In another set of experiments, Newstead et al., (1992) attempted to distinguish between the misinterpreted necessity model and mental models theory, which provides an alternative account of the belief by logic interaction (Oakhill & Johnson-Laird, 1985; Oakhill, Johnson-Laird, & Graham, 1989). Mental model theorists posit that participants do not construct multiple models of syllogisms associated with believable conclusions, but do construct multiple models of syllogisms associated with unbelievable conclusions. This discrepancy has the result of producing a greater belief bias effect for invalid categorical syllogisms than for valid categorical syllogisms and, thus, a belief by logic interaction. Valid syllogisms are associated with a small belief bias effect because their conclusions will be endorsed irrespective of their believability. For example, a valid syllogism's conclusion will be endorsed either because participants agree with it (i.e., it is believable) or because it leads to the construction of multiple models (i.e., it is unbelievable), leading participants to appreciate its validity. Thus, in either case, the conclusion to a valid categorical syllogisms will be accepted. In contrast, for invalid syllogisms, conclusions will be correctly rejected only when they are unbelievable and lead participants to create multiple models of the syllogism. When the invalid syllogism's conclusion is believable, however, participants will incorrectly endorse it (Oakhill & Johnson-Laird, 1985; Oakhill et al.,1989). Newstead et al., (1992) tested two predictions advanced by mental models theorists concerning the belief by logic interaction.

The first prediction tested by Newstead et al., (1992) was that the belief by logic interaction should disappear with single-model invalid categorical syllogisms, irrespective of whether they lead to determinately false or indeterminately false conclusions. The interaction should disappear because only one model of the premises and conclusion needs to be constructed. Hence, the tendency to accept believable but invalid conclusions should decrease since only one model needs to be considered, leaving little room for erroneously accepting an invalid conclusion. In contrast, the misinterpreted necessity model holds that the belief by logic interaction should surface with invalid categorical syllogisms that yield indeterminately false conclusions, regardless of the number of models associated with them. Newstead et al.'s results revealed an absence of the belief by logic interaction with indeterminate, single-model categorical syllogisms, thus supporting mental models theory.

In another experiment, Newstead et al., (1992) tested a second prediction taken from mental models theory. The second prediction was that using multiple-model categorical syllogisms should resurrect the belief by logic interaction since the generation of multiple models in multiple-model syllogisms appears to be mediated by the believability of the conclusions. Results from this experiment indicated a resurrection of the belief by logic interaction. As a final test of the mental models explanation of the belief by logic interaction, Newstead et al., (1992) attempted to nullify the belief by logic interaction by modifying the instructions given to participants. Under this new condition, participants were instructed on the importance of searching for alternate models in spite of the believability of the syllogism's conclusion. Under these modified instructions, the belief by logic interaction failed to appear. In contrast, under the original instructions, the belief by logic interaction surfaced again. From these collective results, Newstead et al., (1992) concluded that mental models theory provided the best approach to understanding belief bias and the belief by logic interaction.

All theories examined so far attempt to explain belief bias and the belief by logic interaction by positing that content influences the manner in which the reasoning task is understood. For example, when an unbelievable conclusion challenges prior beliefs, participants are said to be motivated to carefully inspect the conclusion and either work

out the logic of the argument or search alternate models. In contrast, when the conclusion supports prior beliefs, participants are said to be motivated to accept it without further analysis. Thus, a task's content appears to facilitate or hinder formal reasoning depending on its level of coherence with participants' prior knowledge. Once again, the assumption here is that formal reasoning can be interrupted by ordinary comprehension. It appears, however, that content does not necessarily hinder logical performance, especially if unbelievable content motivates in-depth processing and, consequently, logical performance on invalid and valid syllogisms (Evans et al., 1983).

Inductive Reasoning Processes

What belief bias and the belief by logic interaction suggest is that content via ordinary comprehension processes need not hinder logical reasoning, and may in fact reflect the important role of beliefs in reasoning. It is my assertion that biases observed on abstract and thematic categorical syllogisms are better interpreted as natural responses arising from inductive reasoning processes. When individuals approach reasoning tasks inductively, those tasks that do not fit the constraints of an inductive system will likely elicit inductive behaviours that lead to formal errors. For example, the biases observed on abstract tasks are understandable if one stops to consider that these tasks lack a fundamental input variable to inductive processes–meaningful content. As illustrated in the inductive-coherence model presented in this chapter, inductive processes use prior knowledge to constrain the search for possible solutions to a problem. As a result, the absence of meaningful content hinders the reasoning process, leading participants to rely on simple heuristics in hopes of arriving at any solution. Furthermore, belief bias and the belief by logic interaction are understandable responses given an inductive reasoning system that exploits content in an effort to guide problem-solving. As illustrated in the inductive-coherence model, believable conclusions are not normally analyzed thoroughly because there is little reason to question beliefs that are already known to be true (Evans, Over, & Manktelow, 1993; Harman, 1986, 1995; Quine & Ullian, 1978), unless the purpose of the task includes reasons as to why a known belief should be questioned (e.g., instruction

manipulation, Newstead et al., 1992).

Economy of information processing can be used to explain the belief by logic interaction. Recall that this interaction is illustrated by small effects of belief bias among valid categorical syllogisms but large effects of belief bias among invalid categorical syllogisms. The best explanation for this interaction is one provided by mental models and one that fits well with the inductive-coherence model. Believable conclusions to valid and invalid syllogisms are quickly and correctly accepted because they cohere with what we believe is true about the world and so need not be questioned. In contrast, unbelievable conclusions to valid and invalid syllogisms are analyzed more closely because they present inconsistent information to what we believe to be true about the world so they warrant close examination.

Conditional Syllogisms

Task Definition

Another standard task used to study formal reasoning involves conditional syllogisms. Conditional syllogisms consist of two premises and a conclusion. The first premise involves a conditional rule of the form, "if p then q," where p, the antecedent, and q, the consequent, represent propositions of some kind. The second premise involves one of the propositions (or its negation) represented in the conditional rule, and the conclusion involves the other proposition (or its negation). The structure of a conditional syllogism takes the following form:

<u>If p then q; p</u>
Therefore, q

The conclusion of this conditional syllogism represents the valid inference known as modus ponens (MP). Three additional inferences associated with the conditional syllogism are the valid modus tollens (MT) and the two fallacies, namely, denial of the antecedent (DA) and affirmation of the consequent (AC). These inferences take the following form:

MT: <u>If p then q; not-q</u>	DA: <u>If p then q; not-p</u>	AC: <u>If p then q; q</u>
Therefore, not-p	Therefore, not-q	Therefore, p

As with categorical syllogisms, proof-theoretic and model-theoretic methods exist for evaluating the validity of conditional syllogisms (Oakhill & Garnham, 1994). According to the model-theoretic method, for example, a conditional syllogism such as the one illustrating MP is valid only if the conditional rule is true. The conditional rule is judged to be true either when the antecedent is false or both the antecedent and consequent are true. Given the truth of the conditional rule, MP and MT inferences can be drawn with certainty, whereas DA and AC inferences cannot be drawn with certainty and are, thus, fallacious.

When conditional syllogisms are used in reasoning studies, participants are presented with a series of partially complete syllogisms and typically asked to either generate or choose from a list of alternatives the necessary conclusion. Alternatively, participants are presented with a set of complete syllogisms and asked to indicate whether they accept or reject the conclusion inferred from the premises. Similarly to categorical syllogisms, investigations of conditional reasoning have focused on abstract and thematic conditional syllogisms (e.g., Cummins, 1995; Taplin, 1971).

Biases Observed on Abstract Conditional Syllogisms

The biases observed on conditional syllogisms differ depending on whether abstract or thematic syllogisms are used. With abstract syllogisms, participants typically endorse the fallacious inferences, DA and AC (Braine, 1978; Braine & O'Brien, 1991). Endorsing these inferences is considered biased or incorrect because they do not necessarily follow from the premises. For instance, given "if p then q" and "q" as premises, it is incorrect to infer the conclusion "p" since "not-p" could just as equally be inferred. The conditional rule only specifies that given a "p" a "q" will follow, however, it does not specify that the converse is true. A common explanation of this bias is that participants interpret the conditional rule as a biconditional rule, in which case one can correctly assume the converse of the rule to be true and draw the inferences, DA and AC (Legrenzi, 1969; Taplin & Staudenmayer, 1973).

Although participants normally endorse fallacious inferences (Braine & O'Brien, 1991),

they also correctly endorse valid inferences such as MP and to a lesser extent MT (Braine, 1978). As with categorical syllogisms, mental models theorists and syntactic theorists normally attribute errors on conditional syllogisms to ordinary comprehension processes. For instance, syntactic theorists (Rumain, Connell, and Braine, 1983) note that the reason why almost all children and many adults fail to avoid fallacious inferences is because they do not distinguish between necessary inferences and invited inferences. Necessary inferences result as a consequence of employing analytical comprehension processes, while invited inferences result as a consequence of employing ordinary comprehension processes. Ordinary comprehension processes are presumed to be invoked on conditional tasks because invited, everyday inferences are drawn by participants (e.g., assuming the truth of the converse of the conditional rule). Invited inferences are prohibited in logical tasks.

Because participants correctly endorse MP inferences, syntactic theorists claim this as evidence of formal reasoning processes. Syntactic theorists claim MP inferences involve a straight forward application of a syntactic rule. Furthermore, they explain the lower likelihood of MT inferences by postulating that a sophisticated process called *schema for conditional proof* is needed to draw the inference (Braine, 1978; O'Brien & Braine, 1991; Rips, 1994). According to syntactic theorists, formal reasoning rules exist for drawing valid MP and MT inferences but not for drawing the fallacious DA and AC inferences. Syntactic theorists claim participants endorse fallacious inferences because ordinary comprehension processes interfere with the proper, analytical interpretation of the conditional rule, and not because they possess formal rules for invalid inferences.

In everyday language, the conditional construction "if p then q" has a number of connotations, including promise, obligation, or causal, which suggest a biconditional interpretation (Fillenbaum, 1975). Syntactic theorists claim that participants impose this faulty biconditional interpretation on most conditionals (Braine, 1978; Braine & Rumain, 1983). For example, given the promise "if the lawn is mowed, then you will receive 10 dollars," it is common in everyday language to assume that the converse of this conditional, "if the lawn is not mowed, the you will not receive 10 dollar," also holds;

from a logical perspective, however, the converse need not hold (Fillenbaum, 1975). In essence, people make many unnecessary but practical inferences in logical tasks because such inferences are allowed and, in fact, required in everyday conversation (Grice, 1975).

In contrast to the syntactic approach, mental model theorists propose that people construct a mental model of the conditional syllogism. For example, given the conditional syllogism "if A then B" along with "A," participants construct a model reflecting either a conditional or biconditional interpretation of the conditional rule along with the categorical premise. These two interpretations are as follows:

(a) conditional: [A] B (b) biconditional: [A] [B]

.....

The ellipsis below the conditional interpretation indicates that the model should be expanded to reflect additional interpretations. According to the mental models approach, participants incorrectly endorse the fallacious inferences, DA and AC, because they either do not construct an accurate model initially of the conditional syllogism (e.g., they interpret the conditional biconditionally) or they do not explore alternate models of the conditional due to working memory limitations (Johnson-Laird & Byrne, 1991; Toms, Morris, & Ward, 1993). In the first case, participants may not construct an initially accurate model of the task because they do not appreciate the formal interpretation of a conditional rule, that is, the truth-functional interpretation. Instead, ordinary comprehension processes obstruct the proper conditional interpretation from being modeled and promote modeling a biconditional interpretation (Johnson-Laird & Byrne, 1991). As mentioned previously, a biconditional interpretation of the conditional rule legitimizes the invalid inferences, DA and AC because the converse of the conditional rule can be assumed to be true. In the second case, participants do interpret the conditional rule properly but fail to reject the fallacious inferences or endorse the valid MT inference because they only consider their initial model of the syllogism. That is, they do not construct or search for alternate models that would explicitly show that the antecedent is only sufficient but not necessary for the consequent. In short, proponents of the mental

models approach account for errors on conditional syllogisms by maintaining that participants construct initially inaccurate models of the task or fail to search for alternate models (Johnson-Laird & Byrne, 1991).

In sum, syntactic theorists and mental model theorists attribute fallacious inferences on conditional syllogisms primarily to ordinary comprehension processes and secondarily to working memory limitations. Attributing fallacious inferences to these interfering factors, however, assumes the existence of formal processes. In contrast, an alternate (inductive) explanation promotes both ordinary comprehension processes as well as working memory limitations as constraints inherent to an inductive reasoning system designed to expedite information processing. In particular, although ordinary comprehension processes may fail to promote logically necessary inferences, they do promote pragmatic, experience-based inferences. Moreover, although working memory limitations may limit the number of alternatives (i.e., interpretations of the task) a participant considers to the task, this constraint allows the reasoner to reduce the problem-space and generate quick conclusions (Holland et al., 1986). When ordinary comprehension processes and working memory limitations are viewed as natural constraints arising out of an inductive system, errors on formal tasks need not be viewed as errors in as much as inductive responses to formal tasks.

Biases Observed on Thematic Conditional Syllogisms

Unlike abstract conditional syllogisms, thematic conditional syllogisms use meaningful propositions in both the conditional rule and premise, for example, "if the glass is dropped, then it will break" "the glass is dropped" therefore "it will break" (Staudenmayer, 1975). Thematic conditional syllogisms can be interpreted in a number of different ways as a consequence of linguistic and non-linguistic factors (Cheng & Nisbett, 1993; Fillenbaum, 1975; Staudenmayer, 1975). Linguistic factors include content (e.g., whether the conditional rule and premise are abstract or meaningful), the relationship between antecedent and consequent (e.g., whether the conditional rule reflects nominal class inclusion or causality), and the interpretation a participant imposes on the connective

"if...then" (e.g., whether it is interpreted as "if and only if" or as "because;" see Ormerod, Manktelow, & Jones, 1993 for a discussion of how logically equivalent connectors are interpreted differently). Nonlinguistic factors include pragmatic factors such as participants' empirical knowledge of the world and the ease with which they draw invited or practical but logically unnecessary inferences.

Biases observed on thematic conditional syllogisms are difficult to characterize because, depending on the specific content employed, participants can be induced to infer the fallacies, DA and AC, or to avoid the valid inferences, MP and MT (e.g., Byrne, 1989; Cummins et al., 1991; Cummins, 1995; Digdon, 1986; Elio, 1997; Staudenmayer, 1975). Cummins et al., (1991) demonstrated that the number of "alternatives" associated with both the antecedent and consequent of the conditional rule influences the kind of inferences that are drawn. For example, a causal conditional syllogism that indicates the antecedent is unnecessary for the consequent (e.g., if I eat candy often, then I get cavities) evokes few fallacious inferences from participants but also few valid inferences (Cummins et al., 1991). An antecedent that is unnecessary for producing the effect may represent one of many causes capable of producing the effect. In contrast, a causal conditional syllogism that indicates the antecedent is necessary for the consequent (e.g., If my finger is cut, then it bleeds) evokes higher numbers of valid inferences, but also higher numbers of fallacious inferences (Cummins et al., 1991). Moreover, even if the antecedent is necessary for producing the effect, knowledge of disabling conditions that can prevent the effect from occurring can reduce both fallacies and valid inferences. Consider the following two conditionals:

(a) If my finger is cut, then it bleeds. (b) If I eat candy often, then I have cavities.

In contrast to conditional (a), conditional (b) reflects a situation where many alternative antecedents (or causes) could have produced the consequent in the rule. Furthermore, in contrast to conditional (a), conditional (b) reflects a situation where many disabling conditions could have thwarted the occurrence of the consequent given the antecedent in

the rule. For example, if I eat candy often but I brush equally often, then I should not get cavities. Cummins et al. (1991) provide evidence to suggest that participants accept significantly fewer valid inferences in response to thematic conditionals, which reflect disabling conditions that can prevent the effect from occurring. In contrast, participants accept significantly more valid inferences in response to conditionals that reflect few disabling conditions. According to Cummins et al., (1991), participants who were given conditionals such as (a) accepted valid MP inferences more frequently than those who were given conditionals such as (b). These investigators also show that participants accept significantly more invalid inferences in response to conditionals reflecting few alternative antecedents (i.e., conditional a), whereas they accept significantly few invalid inferences in response to conditionals reflecting many alternative antecedents (i.e., conditional b). In short, depending on the specific content of thematic conditional syllogisms, both valid and invalid inferences may be facilitated or suppressed (see also Byrne, 1989, 1991).

Mental model theorists explain these observations by claiming that the suppression and enhancement of inferences is closely tied to how participants' interpret the premises (Byrne, 1989; Johnson-Laird & Byrne, 1991). In other words, these theorists account for wavering inferences by positing that the kinds of inferences endorsed or rejected rest on participants' interpretation of the situation described in the premises and, ultimately, in the kind of mental model constructed. If the meaningful conditional syllogism biases the participant to construct a model, for example, in which many antecedents can bring about the consequent, then that participant will not likely infer the AC and DA fallacies. Furthermore, if the content of the syllogism biases the participant to construct a model in which many disabling conditions can thwart the effect from occurring in spite of an appropriate antecedent, then that participant will likely fail to make the valid MP and MT inferences. According to mental model theorists, then, participants' ordinary comprehension processes dictate whether participants will reason logically or fallaciously via the model constructed.

Syntactic rule theorists have not been able to satisfactorily account for the suppression of valid inferences because, according to them, valid inferences should not be suppressed

if formal rules trigger them (O'Brien, 1993; Politzer & Braine, 1991). In an attempt to account for the suppression of valid inferences, Politzer and Braine (1991) claim that participants may hesitate to make or endorse valid MP and/or MT inferences within a thematic context if the context sheds doubt on the truth of the conditional rule and premises. These theorists maintain that formal rules can be applied only to premises the reasoner considers to be true. If the context casts doubt on the conditional rule, consequently confusing participants as to the truthfulness of the premise set (Politzer & Braine, 1991; O'Brien, 1993), then valid inferences will be suppressed. This explanation for the suppression of valid inferences allows syntactic theorists to maintain their hypothesis of logical reasoning rules.

The claims made by Politzer and Braine (1991), however, rest on semantic and not on syntactic appeals (Fillenbaum, 1993). The idea that the thematic content of a conditional rule can cause participants to doubt its truth assumes that participants' substantive interpretation of the conditional is important. By implying that the context of the conditional is important to the application of formal rules, syntactic theorists assert semantic, rather than syntactic, explanations of inferential behavior. Moreover, these kind of semantic explanations rely on participants' ordinary comprehension processes in an attempt to account for the invalid inferences made on thematic conditionals (Politzer & Braine, 1991). In sum, both syntactic theorists and mental models theorists rely on ordinary comprehension processes to explain invalid and valid inferential behavior on conditional syllogisms.

Inductive Reasoning Processes

Ordinary comprehension processes are assumed to bias inferential behavior by derailing formal reasoning processes. This assumption rests on the further assumption that formal reasoning processes exist and can, therefore, be derailed. The assumption, however, that formal processes form the basis of human reasoning is only a working hypothesis, and one that seems inadequate given participants' performance on formal tasks and the nature of their environment (Cosmides & Tooby, 1996; Evans et al., 1993). An alternate assumption

is that inductive reasoning processes underlie human reasoning as evidenced by the biases observed on formal tasks and, more importantly, the nature of environment in which human beings must solve problems.

Although mental models theorists and syntactic theorists assume formal processes, they fail to use the formal framework to convincingly account for biases, including the suppression of valid inferences. These theorists fail to convincingly account for biases because they assume ordinary comprehension processes and working memory limitations interfere with inferential behavior without (a) detailing how such interference occurs, and (b) explaining the adaptive nature of such conflicting processes (Chan & Chua, 1994). Moreover, the importance these theorists attribute to ordinary comprehension processes, including participants' degree of belief in the task as well as their background knowledge, suggests that prior knowledge is fundamental to understanding reasoning–a largely inductive proposal (see Chan & Chua, 1994; Elio & Pelletier, 1994; Holland et al., 1986).

Understanding conditional reasoning using an inductive approach seems reasonable given that the conditional construction "if...then" can invoke many different interpretations among participants (Staudenmayer, 1975). However, this does not mean that all interpretations will be fully explored. Although the logical thinker is expected to systematically consider all possible interpretations (i.e., similar to the position advanced by mental models theory) until the best interpretation is reached, this does not imply that people can do so efficiently or that they necessarily should consider all interpretations (Harman, 1986). Working memory limitations may limit the number of interpretations considered, but, more importantly, it may be advantageous for people to consider few interpretations. For instance, Holland et al., (1986) suggest that selective information processing is a useful endogenous constraint in an inductive reasoning system. This constraint, albeit unhelpful in logical tasks, has adaptive benefits because it reduces the problem space and facilitates rapid information processing. As described with the inductive-coherence model, considering multiple interpretations of information does not facilitate a quick and efficient call to action. This is not to say that actions based on a logical analysis of information are maladaptive. In fact, actions resulting from such an

analysis are more likely to be error free and this is highly adaptive, but the issue here is that the time required to fully analyze a problem space conflicts with rapid and economical information processing; a necessity in everyday situations where people sometimes need to think and act fast.

As illustrated with the inductive-coherence model, considering many distinct interpretations can also increase a person's uncertainty about any one interpretation (Garner, 1962; Sloman, 1994). The idea that uncertainty accompanies the consideration of many interpretations has intuitive appeal because there are many everyday situations where increasing the number of options actually reduces our certainty in any one option. For example, many students claim that well-written multiple choice "distractors" cause them to doubt the accuracy of the answer they thought to be correct and would have supplied if the question had been open-ended. Although it is considered logical to consider multiple interpretations of a task, multiple interpretations can decrease a participant's certainty or confidence in any conclusion or action. In sum, if the purpose of everyday reasoning is to solve uncertain problems efficiently, then reasoning processes must process information selectively, consequently reducing the problem space, and increase the reasoner's confidence in the ensuing conclusion or action–a byproduct of selective processing.

Wason's Card Selection Task

Task Definition

As noted in Chapter 1, a task that is famous for its simplicity as well as for the poor logical performance that it elicits from participants is Wason's (1966) card selection task. Wason's selection task is typically viewed as a problem in hypothesis-testing. In standard form, it consists of presenting a participant with an abstract conditional rule, *If there is a vowel on one side of the card, then there is an even number on the other side* (i.e., general form "if p then q"), and four cards displaying instances of a *vowel* (i.e., p), a *consonant* (i.e, not-p), an *even number* (i.e., q), and an *odd number* (i.e., not-q). For instance, Figure 2.4 illustrates how the four cards may be positioned in front of participants. Although

participants see only one side of each card, they are told that the flip side of each card contains information about another category. For example, a *4* or *7* might be found on the flip side of the *A* card, and an *A* or *K* might be found on the flip side of the *4* card. Participants are then instructed to test the truth or falsity of the rule by selecting the fewest possible cards from the set of four.

According to formal proofs, only a pairing of the antecedent (i.e, the "A" in this case) with a negation of the consequent (i.e., the "7" in this case) can falsify and, therefore, test the truth or falsity of Wason's conditional rule (Garnham & Oakhill, 1994). As a result, the *A* card should be turned over in case there is a *7* on its flip side, as well as the *7* card in case there is an *A* on its flip side. Neither the *K* nor the *4* card need to be selected because they cannot falsify the rule but instead only confirm it. For example, even if card *4* were paired with a *K*, it fails to disprove the rule because the rule does not exclude the possibility of an even number following a consonant.

Wason's task appears to be a simple problem, but this simplicity is deceptive; participants invariably and overwhelmingly get it wrong (for a review see Evans, Newstead, & Byrne, 1993). For instance, an average of only 10 percent of participants select the correct cards. In contrast, an average of 90 percent of participants make either incomplete selections by choosing only the *A* card or incorrect selections by choosing the *A* card along with the *4* card. If this task is supposed to be measuring reasoning efficacy, then an average of 90 percent of participants are inefficacious. How does such an inefficacious populace solve problems? There are different positions on this matter.

Biases Observed on Abstract Versions of Wason's Selection Task

When Wason's task is presented in its traditional abstract form, participants overwhelmingly select cards that confirm or match the propositions represented in the conditional rule (e.g., either card "p" or both cards "p" and "q"). Wason (1966) accounted for this bias by concluding that people possessed faulty reasoning processes, which erroneously led participants to verify the rule. In Wason's opinion, little else could justify participants' overwhelming bias to test a conditional rule with confirming evidence

instead of falsifying evidence. Other investigators (e.g., Cox & Griggs, 1983; Evans, 1989; Evans & Lynch, 1973; and Manktelow & Evans, 1979), however, advanced another account of participants' performance. They suggested that instead of confirming the rule with their selection of cards, participants attempted to *match* their card selections to the propositions reflected in the rule. These investigators argued that if participants were operating under a confirmation bias, then given a conditional such as "if p then *not-q*," participants should respond to the rule by selecting the cards "p" and "not-q." In fact, however, these were not the cards participants selected. In studies where the conditional rule was manipulated to include all possible valence combinations (i.e., if p then q, if not-p then q, if p then not-q, and if not-p then not-q), Manktelow and Evans (1979) found that participants displayed a significant tendency to match their card selections to the "topics" or propositions of the rule (regardless of a negation). For instance, given the rule "if p then not-q," participants typically selected cards "p" and "q," suggesting a bias to match their card selections to the topic of the rule rather than to verify the rule (Evans, 1989).

However participants' performance is described, whether by a confirmation bias or a matching bias, there is still no satisfactory explanation as to why participants' performance would be distorted on such a simple task. Perhaps the task is not simple at all. For example, some charge that the task is quite difficult and elicits participants to employ simple heuristics such as confirmation or matching strategies (for a review see Evans et al., 1992). In support of the critics, when the task was modified or simplified, participants' performance did improve dramatically (Evans et al., 1993; Johnson-Laird & Wason, 1970). For example, in the Reduced Array Selection Task (RAST), participants were asked to test the rule, *If they are triangles, then they are black*, by selecting contents from two boxes. One box contained black figures and the other contained white figures, and participants' task was to decide which box needed to be searched in order to test the rule. From a strictly logical perspective, only the box containing the white figures needed to be searched because it alone could potentially contain a triangle (which according to the rule should be black) and falsify the rule. The black box did not need to be searched because its contents were black; that is, the contents could only confirm the rule and, therefore, were

akin to card "q" in Wason's task. Under the RAST conditions, all participants exhaustively searched the box containing the white figures without searching the box containing the black figures. In short, when participants' attention was directed at examining only possible consequents of the rule (i.e., the boxes containing different coloured figures), they correctly searched the box whose contents could falsify the rule. Results from the RAST lend some credibility to notions that Wason's task may be difficult and confuse participants, leading to poor logical performance. Wason (1983), however, disagreed with such critics, claiming that the complexity of the task had been exaggerated to justify poor performance, and eclipse the illogical heuristics that characterize people's reasoning strategies in general.

Evidence that abstract versions of Wason's task may be difficult for participants to solve also comes from a computerized production rule system named PSYCOP (Rips, 1994, 1995). When Wason's task was presented to PSYCOP for a solution, PSYCOP was easily able to deduce card "p" or the antecedent of the rule as a conclusive test of the conditional rule but had difficulty deducing card "not-q" or the negation of the consequent (Rips, 1994). Evidently, PSYCOP deduced the antecedent by taking the conditional rule along with its antecedent as premises and applying an MP rule (i.e., IF elimination rule in Rips' (1994) description). However, the negation of the consequent was not deduced as easily because PSYCOP did not have an MT rule to carry out the inference. PSYCOP overcame this obstacle by assuming the possible conclusions arising from the conditional rule along with different categorical premises and, then, by working back from each conclusion. For instance, given the conditional rule, *if vowel then even*, along with the negation of its consequent, *not-even*, the possible conclusions arising from this pairing are *not vowel* or *vowel*. Once these possible conclusions were generated, PSYCOP used a backward rule to derive the premises that followed from each conclusion. If a conclusion led to a contradiction of the conditional rule (e.g., the conclusion "vowel" from the premises "if vowel then even" and "not-even" contradicts the rule), then PSYCOP tagged the associated categorical premise as providing a conclusive test of the conditional rule. From PSYCOP's performance, Rips (1994) hypothesized that participants might

experience the same difficulty PSYCOP encountered with Wason's selection task; that is, difficulty generating the card "not-q" because it requires not only a more sophisticated application of rules (i.e., forward and backward rules) but the consideration of *possible conclusions* about what lies on each card's flip side.

Mental models theorists offer a similar account of why participants perform poorly on Wason's selection task. Specifically, Johnson-Laird and Byrne (1991) claim that participants who only construct a single model instead of multiple models of the conditional rule are likely to select the propositions reflected in the rule. For example, according to Johnson-Laird and Byrne (1991), construction of a single model of the rule *if there is an A, then there is a 2* will lead participants to focus on the A and the 2 as conclusive tests of the rule. If, however, participants construct alternate models of the rule, ones which show the rule to be true when a letter other than A precedes the 2, but false when a number other than 2 follows the A, then they will appreciate the value of negating the consequent in testing the rule.

In general, formal theorists have had difficulty accounting for the poor performance elicited by Wason's selection task. One source for the difficulty comes from the consideration that if people have trouble falsifying a simple conditional rule via formal rules, how can we possibly expect such formal rules to handle the messy everyday problems people normally have to solve. Rips (1994, 1995) mentions that Wason's task may be difficult because we lack an MT rule and limited working memory capacity to think about possible conclusions, but these should not be obstacles to a system based on logical rules. One would think that possessing the MT rule and the ability to consider multiple conclusions would be central to a system whose function it was to draw logical conclusions. Another difficulty facing formal theorists is the improved performance (i.e., thematic effects) observed on Wason's task when it is framed in a thematic context.

Biases Observed on Thematic Versions of Wason's Selection Task

In contrast to abstract versions, thematic or meaningful content versions of Wason's task commonly elicit from participants what appears to be logical performance. For

example, given a thematic version of the task, participants are much more inclined to select the cards "p" and "not-q" in response to the conditional rule (Evans, 1989; Evans et al., 1993). Depending on the specific content or context[1] used in thematic versions, some investigators report that up to 78 percent of participants exhibit logically correct performance on the task (e.g., Cosmides, 1989). Formal theorists such as Wason (1983; Griggs, 1983; O'Brien, 1993) have argued, however, that most thematic versions of the task do not measure logical reasoning but instead measure something else; for instance, memory of the content or context described in the task.

A number of theories have been proposed to explain thematic effects on Wason's task. Cheng and Holyoak (1985, 1986), for example, in their pragmatic reasoning theory[2] maintain that most thematic versions of the task, which facilitate performance, involve a "permission" context where pre-conditions and actions are negotiated. This permission context elicits a permission schema in participants that guides their responses and actions in such a situation. In particular, two out of the four production rules that make up the schema, *if the action is to be taken, then the precondition must be satisfied* and *if the pre-condition is not satisfied, then the action must not be taken*, orient participants to select the correct cards in response to the task. Cheng and Holyoak (1985) show the appropriateness of the permission schema by using it to explain performance on the "drinking problem" (Griggs & Cox, 1982), a thematic version known to reliably facilitate

---

[1]Content is differentiated from context in that the former refers to the theme of the task, whereas the latter refers to the specific scenario or framework of the task (Pollard & Evans, 1987). For instance, the content of the "drinking problem," a commonly used thematic version of Wason's task, is about restrictions on who may drink alcohol, but its context is about an enforcer of a rule who checks for potential violations.

[2]Theories similar to Cheng and Holyoak's (1985) pragmatic reasoning theory include Cosmides' (1989) social contract theory and Gigerenzer and Hug's (1994) cheating detection theory. According to social contract theory, individuals possess a social contract algorithm developed out of an evolutionary necessity for "adaptive cooperation between two or more individuals for mutual benefit" (p.193). Specifically, the algorithm is induced in those situations assuming a cost-benefit structure, that is, *if you take the benefit, then you must pay the cost*. Cheating detection theory is an extension of social contract theory that includes how different perspectives influence what is assumed to be a benefit or a cost.

performance.

In the "drinking problem," participants are presented with the conditional rule, *if a person is drinking beer, then the person must be over 19 years old,* along with the cards, *beer*, *coke*, *over 19 years*, and *under 19 years* (Cox & Griggs, 1982). Participants are instructed to pretend that they are enforcers of the rule, and to select those cards that represent potential violations of the rule. Cheng and Holyoak (1985) interpret this well-known problem as a classic permission situation, in which drinking beer is seen as a desirable action and being over 19 years of age is the pre-requisite condition to carry out the action. According to Cheng and Holyoak (1985), the schema's propositional rules, *if the action is to be taken, then the precondition must be satisfied* and *if the pre-condition is not satisfied, then the action must not be taken*, guide the selection of cards by highlighting the cases when the rule is violated (e.g., if he is under 19 years old, then the person cannot be drinking beer). Cheng and Holyoak maintain the permission schema and others like it (e.g., obligation and causal schema, see Cheng & Nisbett, 1993) arise inductively through people's environmental experiences and their information processing mechanisms, which function to structure both the environment and responses to it.

Although schemas are important descriptors of human behaviour, Manktelow and Over (1991) note that Cheng and Holyoak's permission schema may be better conceived as an undeveloped collection of *deontic* rules, which are invoked in situations calling for deontic reasoning. They describe deontic reasoning as taking place when,

> [W]e consider what we may (or are allowed to) do, or what we ought to (must or should) do, rather than what was, is, or will actually be the case. So basic forms of deontic thought are concerned with permission and obligations. (p.88)

Deontic reasoning is enabled by subtle considerations of semantic, pragmatic, and social information that influence a person's assessment of the utilities of possible actions. Manktelow and Over (1991) suggest that while Cheng and Holyoak's permission schema is deontic in character, it fails to specify how individuals assess utilities; a variable considered fundamental to deontic thinking. Furthermore, they point out that the productions rules that make up the permission schema, themselves incorporate deontic

terms (e.g., may and must) that need to be deconstructed by a more basic "deontic" schema. In sum, these investigators question the need for "middle of the road" schemas whose resulting effects might be explained by more basic schemas.

In spite of the appeal of pragmatic reasoning theory and others like it, some investigators point out that some fundamental variables have been left unclear, such as how participants interpret or understand the information presented in the task (Chan & Chua, 1994; Liberman & Klar, 1996; Manktelow & Over, 1991; Pollard & Evans, 1987). In an effort to untangle this ambiguity, Liberman and Klar (1996) suggest that participants approach the task just as they approach any other everyday reasoning problem and, as such, task comprehension is critical to performance. These investigators suggest that further study needs to be devoted to how (a) conditional rules are interpreted in thematic versions of Wason's selection task, (b) intelligible the context of the task is for participants, and (c) appropriate a falsification strategy is given the context of the problem (Klayman & Ha, 1987). The latter point is especially germane if it is assumed that participants approach thematic versions of Wason's task as they would other everyday reasoning problems. In everyday problems, whether one focuses on positive or negative instances with which to test a hypothesis depends on the relative efficiency of testing (and finding) positive versus negative instances. Searching for positive instances to test a hypothesis, for example, is appropriate if the likelihood of finding negative instances is small or the population represented by "not-q" is comparatively larger than "p" or "q" (Klayman & Ha, 1987). For example, if one wanted to test the hypothesis, *if somebody eats spoiled food, then he or she becomes ill*, it makes little sense to test this hypothesis by finding healthy people (i.e., not-q) because one is unlikely to find healthy people who have eaten spoiled food. The population of healthy people is much larger than either those who have eaten spoiled food (i.e., p) or are ill (i.e., "q") (Liberman & Klar, 1996). It makes more sense to restrict the search to those individuals who are ill (i.e., q) because one is more likely to find an ill person who has also eaten spoiled food (i.e., p).

Whether one focuses on positive or negative instances with which to test a hypothesis also depends on a person's goal or "mental set," and the maturity of the hypothesis under

study (see Oaksford & Chater, 1994, for a similar argument but from the perspective of optimal data selection). For instance, a hypothesis considered by participants to be new or not well established may elicit a confirmation strategy over a falsification strategy if participants feel compelled to first establish its existence. In sum, Liberman and Klar (1996) suggest that comprehension factors need to be examined as diligently as the facilitative effects of some contents because a thorough understanding of how individuals interpret these tasks may forfeit the need to posit highly specialized schemas or algorithms. While pragmatic reasoning theory and other similar theories (e.g., Cosmides' (1989) social contract theory and Gigerenzer and Hug's (1994) cheater detection theory) might be overstating the necessity for highly specialized schemas for domain-specific reasoning, they have certainly enriched the dialogue used to discuss reasoning processes with their introduction of a more inductive vocabulary. The vocabulary is no longer restricted to a description of premises and logical rules, but now includes at least reference to participants' goals and beliefs about the task.

Highly specialized algorithms are not the only accounts of thematic effects, however. Mental model theorists, for example, account for thematic effects in much the same way as they account for performance on abstract versions of Wason's task. In particular, they claim that thematic versions of Wason's task facilitate performance when the content promotes the construction of alternate representations (models) of the conditional rule. Syntactic theorists offer no real account of thematic effects except to say that content may facilitate the application of rules (Rips, 1994).

In sum, pragmatic reasoning theory and other similar theories have developed in large part as a response to the thematic effects observed on some meaningful versions of Wason's task. According to these theorists, thematic effects reflect the domain-specific nature of human reasoning. In contrast, supporters of syntactic and mental model theory view thematic effects as products from general inferential processes that might be facilitated when the task is framed in a meaningful content. Specifically, many syntactic reasoning theorists (e.g., Griggs, 1983; Rips, 1995; Wason, 1983) acknowledge that thematic versions of Wason's task lead participants to improved performance, although

they question the exact reason for the facilitation. One appraisal is that thematic versions do not measure logical reasoning, as is assumed of abstract versions of Wason's task, because they require participants to select violations of a rule instead of conclusive tests of a rule's truthfulness (e.g., Liberman & Klar, 1996). In particular, critics argue that "violation" instructions induce participants to think of counterexamples to the rule without requiring them to appreciate the logical structure of the task (i.e., Popper's hypothetico-deductive style of hypothesis testing). Hence, participants may search for violations to the rule without knowing anything about testing its truth. In reality, the processes by which thematic versions improve performance are still under debate given that the effects are sometimes unreliable, that is, not all thematic versions lead to improved performance (Cox & Griggs, 1982; Griggs, 1983; Johnson-Laird & Byrne, 1991; Liberman & Klar, 1996; Maktelow & Evans, 1979; Pollard & Evans, 1987). The processes by which thematic effects are produced may include anything from prior memory of violations of a rule (i.e., Manktelow's & Evans' (1979) *memory-cuing hypothesis*; see also Griggs, 1983) to domain-specific schemas associated with reasoning in particular contexts (e.g., Cheng & Holyoak's pragmatic reasoning theory or Cosmides' social contract theory). In sum, despite the number of theories proposed to explain thematic effects, no theory has been able to fully anticipate the themes that will facilitate performance and, therefore, no theory has unequivocally articulated the exact conditions that promote these effects.

Inductive Reasoning Processes

Findings obtained from abstract and thematic versions of Wason's task reflect the powerful influence of content upon reasoning. Put another way, the findings reflect the intimate and facilitative interaction between knowledge and reasoning. What do these results suggest about human reasoning? According to Cheng and Holyoak (1985, 1989), the results suggest that human reasoning was not designed to deal effectively with problems devoid of content, but was designed to deal with specific content areas. Although their strong domain-specific view of reasoning remains to be conclusively ascertained, Cheng and Holyoak's (1985, 1989) focus on the critical role content plays in

reasoning is significant in the move toward viewing reasoning on formal tasks as originating from inductive processes. As pointed out earlier in the inductive-coherence model, a task's content elicits background knowledge, which in turn facilitates reasoning and problem solving by imposing relevant information on the problem at hand (see also Holland et al., 1986).

The facilitative effect of knowledge, however, does not necessarily imply a facilitation of *logical* reasoning or the guarantee that an absolutely best solution will be found. In this respect, knowledge can either facilitate or hinder logical performance. For instance, thematic versions of Wason's task generally facilitate performance but recall that thematic categorical syllogisms can hinder performance (i.e., belief bias effects). In the latter case, the believability of a syllogism's conclusions can invoke a "mental set," wherein participants seize upon one interpretation of the conclusion and forestall searching for alternate interpretations (Evans, 1989).

The role of content in reasoning is complex. Whereas from an inductive perspective knowledge is expected to facilitate reasoning because it decreases the problem space and expedites information processing, knowledge need not expedite logically correct solutions. The solution generated must be evaluated against its utility for the reasoner and such a solution may not always be logical according to normative standards (Cohen, 1981). In short, although background knowledge might not facilitate logical reasoning per se, it does facilitate effective solutions to everyday problems. Understanding the kind of inductive, everyday reasoning that people employ to solve everyday problems may explain participants' errors on formal tasks. The features of this kind of reasoning are explored next.

Everyday Tasks, Background Knowledge, Non-monotonicity, and Why We Take the Path of Least Resistance

In *Human Problem Solving* (1972), Newell and Simon allege that "to the extent that [human] behavior departs from perfect rationality, we gain information about the psychology from the subject, [and] about the nature of the internal mechanisms that are

limiting his performance." From the preceding review of the literature, it is clear that people generally do not perform logically on formal tasks; that is, people depart from perfect rationality. A central idea developed in this chapter is that people's poor logical performance should not surprise us, as it has in the past (e.g., Begg & Harris, 1982; Braine, 1978; Ceraso & Provitera, 1971; Downing, Sternberg, & Ross, 1985; Janis & Frick, 1943; Johnson-Laird & Tagart, 1969; Kareev & Halberstadt, 1993; Lefford, 1946; and Rips, 1995), if the theoretical framework for understanding human reasoning takes account of the (everyday) problems our reasoning system has adapted to solve. Viewed in this way, poor performance on formal tasks may indicate that the (formal) internal mechanisms we have heretofore assumed of human reasoning are incorrect and that alternate mechanisms need to be considered. Indeed, individuals' poor performance may even begin to appear rational once we view them using an informal, inductive light (Harman, 1986, 1995).

As mentioned previously, interpreting formal performance by way of inductive processes is not a new idea. A number of investigators have either hinted at or fully advanced this idea (e.g., Barsalou & Goldstone, 1998; Cosmides & Tooby, 1996; Evan et al., 1993; Liberman & Klar, 1996; Oaksford & Chater, 1994; Staudenmayer, 1975; Sloman & Rips, 1998; Stevenson & Over, 1995). Some investigators have developed influential reasoning theories based *implicitly* on the inductive treatment of formal tasks. At the root of these theories, however, *formal* reasoning errors are *still* explained in traditional formal terms, viewing them as deviations arising from the existence of interfering factors (e.g., Cheng & Holyoak, 1985; Johnson-Laird & Byrne, 1991). Consequently, these alternative theories represent neither distinct alternatives to the formal approach (i.e., assuming the existence of a logical, formal architecture) nor do they fully advance an inductive approach to reasoning performance. It is for this reason that an explicitly inductive account of formal performance needs to be offered. I have proposed the inductive-coherence model, but there may be others. The strengths of the inductive-coherence model are that it incorporates important inductive features such as background knowledge and beliefs, ill-defined environmental tasks, and the role of consistency or

coherence between beliefs and solutions. The philosophy and concepts from which the inductive-coherence model originated are described below.

Beliefs and Coherence Theory: Building Confidence

Background knowledge and beliefs are critical to solving everyday problems because they supply the reasoner with the information to solve ill-defined problems. Unlike deduction, induction cannot be understood as an "independent, logical system that is largely independent of one's knowledge of the world" (Bisanz et al., 1994, p.183). Knowledge and beliefs are critical to inductive processes so some of their attributes need to be addressed. In particular, consistency among beliefs is fundamental to reasoners if they are to (a) evaluate arguments against their own system of beliefs, (b) build theories to help them anticipate environmental events, and (c) plan and meet goals (Quine & Ullian, 1978; Rescher, 1980; Thagard, 1989). Consistency among beliefs allows reasoners to note when available evidence supports or contradicts their overall belief system (Quine & Ullian, 1978). Some philosophers argue that consistency among beliefs forms the basis of human rationality since without consistency it would be impossible to imagine making sense of the world (Harman, 1986, 1995; Quine & Ullian, 1978).

Consistency among beliefs, however, is not the same thing as reasoning about beliefs. Harman (1986, 1995) notes that reasoning refers to a psychological process that can lead to *possible changes* of beliefs or plans, whereas consistency refers to the *relation* among beliefs. The distinction is important because, as Harman (1986) points out, they are associated with different consequences. For example, the propositions *I am afraid of the dentist* and *dentists use sharp instruments* may be consistent with one another, but this does not justify inferring *dentists use sharp instruments* from *I am afraid of the dentist*. Furthermore, a set of propositions may imply an additional proposition, but if one has little interest in drawing this additional proposition then it should not be drawn in order to avoid unnecessary clutter. Maintaining consistency among beliefs is one way in which Harman (1986, 1995) believes that deductive logic is relevant to a theory of rationality because it can be used to maintain consistent relations among beliefs (Harman, 1986,

1995; Quine & Ullian, 1978).

Given the importance of maintaining consistency among our beliefs, some theorists propose that reasoning or belief revision would be well described using *coherence theory* (Gärdenfors, 1992; Thagard, 1989). Coherence theorists suggest that one's current beliefs are justified "just as they are in the absence of special reasons to change them, where changes are allowed only to the extent that they yield sufficient increases in coherence" Harman, 1986, p.32). Coherence theory embodies a *conservative* approach to belief revision because only minimal changes are supported in an effort to achieve consistency among beliefs (Gärdenfors, 1992). Determining whether a conclusion or argument should be endorsed according to its level of coherence with background beliefs is reflective of the important role background knowledge holds in reasoning (Holland et al., 1986). Such conservatism in belief revision might be viewed as practical since it would be inefficient for a reasoner, intent on learning and increasing his or her knowledge, to casually throw away a body of beliefs that has emerged from direct experience and that has proven useful in the past.

Using coherence theory to conceptualize reasoning might provide clues as to why reasoners view some logical arguments as psychologically stronger than others and why they tend to endorse (or confirm) arguments that coincide with their pre-existing beliefs. First, people might automatically endorse believable categorical conclusions when responding to categorical syllogisms (i.e., the belief bias effect) (e.g., Newstead et al., 1992) because there is little need to doubt or question conclusions already known to be true. Second, people might be more likely to infer particular conditional conclusions when they perceive a specific relationship between the antecedent and the consequent of the conditional rule, suggesting that such conclusions are highly likely. For example, people are much more likely to fallaciously conclude "p" given the premise set "If p then q, and q," if the relationship between p and q is fairly strong. In contrast, individuals are much less likely to commit this fallacy if the relationship between "p" and "q" is weak because there are hypothetically many other antecedents capable of producing "q" (e.g., Cummins, 1995; Cummins et al., 1991). The perceived strength, moreover, between "p" and "q" is

likely learned from a person's environment and the associations present therein (Cummins et al., 1991; Elio, 1997).

Third, recall that people also make the mistake of seeking confirmatory evidence to test a conditional rule instead of seeking falsifying evidence. This tendency, most often observed on abstract versions of Wason's task, is interpreted by many investigators as evidence that people incorrectly assume the truth of the conditional rule (see Evans et al., 1993 for a review). However, from the perspective of coherence theory, it is reasonable to assume the truth of the conditional given that conditional rules constructed in everyday circumstances are *typically* true and do not often need to be falsified. Faced with a rule that is likely to be true, why would one aim to test it by seeking disconfirming evidence where none might exist, when one could more easily test it by seeking confirming evidence (Klayman & Ha, 1987; Liberman & Klar, 1996)? In short, the bias to draw, endorse, and test conclusions that correspond to our beliefs might persist because maintaining consistency among our beliefs is fundamental to human cognition, and also because existing beliefs provide useful information about the state of the world (Bendixsen, Dunkle, & Schraw, 1994; Swain, 1979; Thagard, 1989).

To be sure, an inconsistency between an argument and an individual's existing beliefs does not always imply that the person will disregard the argument. Individuals have two options in this case. They can either trust their beliefs and choose to reject the argument or accommodate the argument by revising some of the beliefs that conflict with it (Thagard, 1989). Both these options function to maintain consistency. Often times people may prefer to disbelieve a conclusion or argument if believing it implies that a series of their existing beliefs need to be thrown out. Unfortunately, the conservative nature of belief revision can hinder the acceptance of a new (correct) belief if accepting it requires forfeiting entrenched pre-existing beliefs. This conservative process, however, is prudent if it is compared to the analogous behaviour of scientists; scientists do not often dismiss a useful theory simply because of one discrepant empirical finding. Throwing away established theories in light of a single challenge or wayward observation would be impractical. If coherence theory describes human reasoning, then the "confirmatory" biases traditionally

observed on formal tasks should not be unexpected (Harman, 1986, 1995).

Another hypothesis arising out of coherence theory is that an argument's believability will be proportional to the consistency of the available evidence. For example, Sloman (1994) suggests that if an argument's premise evokes a belief that conflicts with the belief evoked by its conclusion, a person's confidence in the whole argument is diminished. Sloman (1994) found that when different underlying explanations were attached to the premise and conclusion, people's overall confidence in the argument as measured by a subjective probability scale was reduced. In contrast, he found that when similar explanations were attached to both the premise and conclusion, people indicated greater confidence in the argument. According to coherence theory, when the premise and conclusion share a common explanation, the evidence for the argument increases as well as our confidence in it (Thagard, 1989).

Unlike Sloman (1994), Elio (1997; also see Elio & Pelletier, 1994) challenged participants' confidence in arguments not by invoking different explanations for the premise and conclusion, but by directly challenging the truth of an argument's conclusion. For example, Elio (1997) gave participants the following conditional argument whose conclusion was later contradicted:

> If water was poured on the campfire, then the campfire went out.
> Water was poured on the campfire.
> Therefore, one believes that the campfire went out.

After reading the argument, participants were told that the conclusion did not hold; that is, the campfire had not gone out. Participants were then asked to assume this new conclusion and to indicate their degree of belief in the premises of the argument so as to identify the proposition that most likely led to the contradicted conclusion. Results from this study indicated that when the contradicted conclusion was deduced from a causal conditional associated with few variables that could interfere with its consequent (i.e., few disablers), participants indicated a higher degree of belief in the conditional rule than in the

categorical premise. In contrast, when the contradicted conclusion was deduced from a causal conditional associated with many variables that could interfere with its consequent (i.e., many disablers), participants indicated a higher degree of belief in the categorical premise than in the conditional rule. This trend was primarily observed with MP inferences. Although a similar trend was observed with MT inferences, participants in general indicated a lower degree of belief in the conditional rule than in the categorical premise (Elio, 1997).

One reason why participants might express greater belief in causal conditionals with few disablers than in causal conditionals with many disablers is because they recognize the differences in frequency associated with different antecedents and consequents; that is, they recognize the frequency with which the "consequent is entailed by the antecedent, in all (plausible) worlds a reasoner can generate" (Cosmides & Tooby, 1996, p.6). Elio (1997) argues that these results support "the view that belief revisions are a function of knowledge about the relations between the antecedent and consequent" (p.5). Likewise, others such as Collins and Michalski (1989) argue that the *multiplicity* or the number of possible alternatives associated with the argument and/or referent influences the kind of inferences that we make.

In short, reasoning as evidenced by performance on typical deductive tasks and belief revision tasks appears to be strongly influenced by a need for coherence. Coherence is achieved by either accepting propositions or arguments that support one's existing belief system or by rejecting propositions or arguments that challenge it without sufficient and compelling justification. Moreover, multiple alternative explanations for a proposition, all of which are inconsistent with each other, can function to decrease a person's degree of belief in the proposition. In cases where multiple explanations for a proposition or argument render it uncertain, a person might forego a definite judgement until more evidence is obtained (Quine & Ullian, 1978).

Formal Processes: Considering Alternatives, Ordinary Comprehension Processes, and Suppression Effects

Although the consideration of alternative solutions can decrease a person's degree of belief in any one solution, considering alternatives is critical to logical reasoning (e.g., Braine & Rumain, 1983; Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991; Markovits, 1984, 1985; Rissland, 1986). In fact, in Jean Piaget's final stage of cognitive growth, the formal operational stage, he suggests that adolescents should be able to systematically test (all possible) hypotheses in a problem space (Piaget, 1977). The consideration of alternatives appears to lead to a paradox because it is necessary for logical thought and yet also seems to decrease a person's confidence in the solution. Two points should be noted with respect to this paradox. First, the consideration of alternatives appears to decrease confidence only when the alternatives are inconsistent with each other (e.g., Sloman, 1994). Second, the consideration of alternative solutions appears to only arise when a person is not immediately aware of a solution and therefore consults different options. As pointed out in the inductive-coherence model of reasoning presented earlier, when a person is immediately comfortable with a solution, considering alternate solutions is not pursued and the person conforms to reasoning according to the path of least resistance.

The weight of the empirical evidence suggests that considering a single alternative or reasoning via the path of least resistance is not likely to yield logical responses (Evans et al., 1993). To this end, making alternate perspectives salient to participants for the purpose of improving logical performance is essential. For example, Wason (1977), found that when participants in an experimental group were made aware of contradictory inferences about employees' ages and their salaries, they made fewer fallacious inferences than participants who were left unaware. In general, it appears that increasing participants' awareness of alternate antecedents (or alternate causes as Elio (1997) calls them) reduces their tendency to draw fallacious inferences.

Working from the syntactic or mental logic approach, Rumain et al., (1983) have also demonstrated that people are less likely to draw invalid conditional inferences when they are explicitly told that the consequent in question could have arisen from alternative

antecedents. These investigators, however, conclude that the tendency to draw fallacious inferences grows out of ordinary comprehension processes rather than a misapplication of reasoning rules since the fallacious inferences are minimized with explicit instructions. In another example, Markovits (1984) presented two groups of participants with a story about a boy named "David," who got into bad mood when he had homework to do. One group of participants was instructed to imagine a number of other possible reasons why "David" might get into a bad mood, while the other group was not given these instructions. Afterwards, participants from both groups were asked to evaluate a series of abstract formal inferences. Results from this study indicated that those participants who were both instructed and able to generate a number of alternate reasons for David's bad mood were more accurate than the control group in their evaluations of formal inferences (see also Markovits, 1985, for additional studies suggesting that awareness of alternate antecedents is associated with facilitated logical performance). Although the formal inferences were not related to David's story, Markovits' findings suggest that considering alternate antecedents in general facilitated logical performance.

Rumain et al., (1983) and Markovits (1984, 1985) conclude that encouraging participants to consider alternate antecedents to a conditional statement facilitates *logical reasoning*. Specifically, Markovits (1985) states that "formal competence in conditional reasoning problems may require the capacity to generate and (reason on) a complex network including the given $p \rightarrow q$ relation and other, unspecified but possible, relations of the form $\sim p \rightarrow q$" (p.241). According to this statement, ordinary comprehension processes play a pivotal role in the drawing of fallacious inferences because they mediate a reasoner's acknowledgement of alternate antecedents. Although these theorists claim that ordinary comprehension processes affect fallacious inferences, they need to clarify (a) why invalid inferences can be *hindered* or facilitated depending on the kind of semantic (background) information presented (e.g., imagining alternate antecedents is sometimes facilitated by content), and (b) why valid inferences, which are not supposed to be affected by semantic information, can be suppressed within some contents.

This latter point is important because syntactic theorists claim that valid inferences

cannot be influenced by content since they are drawn by formal rules (e.g., Braine, 1978; Braine & Rumain, 1983; Rips, 1994). However, a number of studies reveal that valid inferences can be suppressed as easily as fallacious inferences when participants consider alternative consequents or disabling conditions that can impede the consequent mentioned in the rule from occurring (e.g., Byrne, 1989; Cummins et al., 1991)). In one study, for example, Cummins et al. (1991) asked participants to rate the following conditional arguments:

(a) If I eat candy often, then I have cavities.
<u>I eat candy often.</u>
Therefore, I have cavities.

(b) If my finger is cut, then it bleeds.
<u>My finger is cut</u>.
Therefore, it bleeds.

Argument (a)'s conditional rule easily evokes alternate antecedents since cavities may result for reasons other than eating candy often; cavities may result because of improper dental hygiene or failing to go to the dentist. In contrast, argument (b)'s conditional rule does not easily evoke alternate antecedents since it is difficult to imagine another cause for a bleeding finger. Not surprisingly, Cummins et al., (1991) found that participants endorsed markedly fewer fallacious inferences on arguments that evoked alternate antecedents (i.e., argument a). More importantly, however, and in opposition to predictions made by syntactic theorists, they also found that participants endorsed markedly fewer *valid* inferences on arguments that evoked alternate consequents (or disabling conditions that might impede the consequent mentioned in the rule from occurring); that is, arguments whose antecedent was *insufficient* to yield the consequent. Argument (a)'s conditional rule evokes alternate consequents because cavities do not necessarily result from eating candy often. If one brushes after eating candy, cavities may not occur. In contrast, argument (b)'s conditional rule fails to evoke alternate consequents. Few consequences, other than bleeding, are likely to occur after cutting one's finger.

In later work, Cummins (1995) proposed that both alternative antecedents and

consequents mediate reasoning by influencing how reasoners conceptualize the causal necessity or sufficiency of a rule's antecedent for bringing about the consequent. Specifically, Cummins (1995) suggested the following:

> [A]lternative causes cast doubt on the necessity of the cause in question for bringing about the effect, whereas disabling conditions cast doubt on the sufficiency of the cause to bring about the effect. The evaluation of necessity and sufficiency relations produces the entailment relations that characterize *naive causal deduction*. (italics added) (p.647).

Naive causal deduction could also be termed *induction* because an inductive system exploits the contextual information (and background knowledge) needed to consider alternatives. However, not any context will suffice to facilitate the consideration of alternatives. The context has to be rich in information in order to cue the consideration of multiple alternatives. As pointed out in the inductive-coherence model, if the context is not rich or the reasoner is highly confident about a single approach, the consideration of alternatives may not occur (Holland et al., 1986).

Following Cummins (1995; and Cummins et al., 1991), but working under the mental models approach, Byrne (1989; see also Johnson-Laird & Byrne, 1991) also found that contextual information can suppress participants' valid inferences. For example, Byrne (1989) presented participants with one of three conditional arguments: (a) simple conditional, (b) simple conditional along with information about an *alternate* antecedent, and (c) simple conditional along with information about an *additional* antecedent. The following is an example of a simple conditional with an additional antecedent:

(Original conditional)	If she has an essay to write, then she will study late in the library,
(Additional conditional)	If the library stays open, then she will study late in the library,
(Categorical premise)	<u>She has an essay to write</u>.
	Therefore, ?

Participants in each of the three conditions were asked to assess both valid and fallacious

arguments. From this manipulation, Byrne (1989) found that participants who assessed simple conditionals together with information about an alternate antecedent made fewer fallacious inferences than participants in the other conditions. In contrast, Byrne found that participants who assessed simple conditionals together with information about an additional antecedent made fewer *valid inferences* than participants in the other conditions. From her results, Byrne hypothesized that information about an additional antecedent cast doubt on the *sufficiency* of the antecedent of the original conditional to bring about the consequent (this is similar to the concept of disabling conditions (Cummins et al., 1991), in which there is little confidence in the sufficiency of the antecedent for bringing about the consequent).

Byrne tested this hypothesis by incorporating a categorical premise that included both the antecedent from the simple conditional and additional conditional as follows:

(Original conditional)	If she has an essay to write, then she will study late in the library,
(Additional conditional)	If the library stays open, then she will study late in the library,
(Categorical premise)	<u>She has an essay to write *and the library stays open*</u>.
	Therefore, she will study late in the library.

Under this manipulation, participants' valid inferences increased without an indication of a suppression effect. From these findings, Byrne (1989) concluded that the suppression of both valid and invalid inferences is closely tied to people's comprehension of the premises. Byrne concluded this because the suppression effects observed depended on whether participants were presented with an "alternate" conditional (i.e., the second conditional rule presents information about an alternate antecedent) or an "additional" conditional. If participants understood the antecedent from the second conditional as an additional requirement for the consequent to occur, then MP and MT inferences were suppressed with the presentation of only one antecedent ("p") in the categorical premise (or "not-q"). In contrast, if participants understood the second antecedent as an alternate requirement for the consequent to occur, then AC and DA inferences were suppressed with the

presentation of "not-p" or "q" in the categorical premise. Moreover, she concluded that mental model theory is better able than syntactic theory to accommodate these findings since "the process of interpretation has been fairly neglected in the inferential machinery proposed by current theories based on formal rules. It plays a more central part, however, in theories based on models" (p.79).

In defence of the syntactic approach, Politzer and Braine (1991) have questioned Byrne's interpretation of the suppression of valid inferences. As mentioned previously, Politzer and Braine argued that one reason why participants likely failed to draw the valid inference in Byrne's third condition was because the additional conditional cast doubt upon the original conditional rule. In retort, Byrne (1991) claimed that participants could not have been confused about the truth of the original conditional since they drew the valid MP inference when the categorical premise included *both* the original and the additional antecedent.

From Cummins' (1995; also Cummins et al., 1991) and Byrne's (1989, 1991; see also Stevenson & Over, 1995) studies, there is compelling evidence that people alter both valid and invalid inferences depending on the context of the argument, which suggests the inductive and non-monotonic nature of human reasoning. Syntactic theorists have had difficulty accounting for these effects as observed by Politzer's and Braine's (1991) critique of Byrne's conclusions. Politzer and Braine's (1991) critiques rest primarily on semantic rather than on syntactic grounds (Fillenbaum, 1993). The difficulty lies in that results repeatedly reveal that how an individual interprets the context of the task will have a significant influence on his or her solution (Byrne, 1989, 1991; Cummins et al., 1991; Evans et al, 1993; see Galotti et al., 1986 for evidence that rules and mental models are employed when reasoning on formal tasks). In fact, Politzer and Braine's (1991) critique might be viewed as an inherently inductive explanation of suppression effects. Indeed, Rumain et al.'s (1983) as well as Markovits' (1984, 1985) results might also be viewed as evidence for inductive processes. Regardless of whether the consideration of alternatives encourages an individual to (a) recall prior knowledge about the frequency of the antecedent and consequent in a rule or (b) doubt the truthfulness of the relationship

between the antecedent and consequent, the inferences remain tied to the context of the rule. In short, the consideration of alternatives seems to mediate logical performance by influencing the reasoner's confidence or trust in the relationship set out by the rule (see also Elio, 1997) and this process is better understood from an informal, inductive approach than from a formal one.

Finally, an inductive account of reasoning would not be complete without mentioning the role that goals play in guiding inferential behaviour (Harman, 1986, 1995). Goals are important to consider because, just as background knowledge, they facilitate reasoning by minimizing the problem space; only solutions that bring a reasoner closer to his or her goals are considered.

The Purpose of Reasoning: Goal-directed Activity

From an adaptive perspective, individuals must learn how to predict events in their environment in order to avoid negative outcomes and to promote positive outcomes. Goals associated with avoiding negative outcomes while seeking positive outcomes are fundamentally based on an individual's beliefs about utility (Fodor, 1981; Harman, 1986, 1995; Holland et al., 1986). The necessity of having and meeting goals has been neglected by formal approaches to reasoning. Mental model theorists also make the implicit assumption that model construction is motivated by reasoners' goals but this assumption is often neglected when reasoners' models are evaluated against the logic of the problem and not against their goals.

In most formal reasoning tasks, the participant is asked to decide which is the best conclusion to draw given a set of premises (e.g., conditional and categorical syllogisms) or to decide which is the best evidence with which to test a statement (e.g., Wason's selection task). Investigators in the past did not expect that a reasoner's goals should interfere with performance on these tasks because they assumed that formal reasoning was not mediated by informal factors, such as strength of belief in the likelihood of the premises and/or conclusion or strength of belief in the utility of the inference. However, there is compelling evidence to indicate that beliefs play a notable role in the product of

formal reasoning (e.g., Cummings et al., 1991; Cummings, 1995; Elio, 1997; Sloman, 1994; Stevenson & Over, 1995). In many influential theories, however, the role of belief is commonly relegated to an interfering factor that biases reasoning.

Some investigators believe that normative models of reasoning misrepresent the "bias" in reasoning (e.g., Einhorn & Hogarth, 1981; Evans, Over, & Manktelow, 1993). These investigators suggest that individuals may impose subjective values or beliefs in their reasoning in order to achieve an end result that is consistent with their goals and values. In their efforts to obtain "subjective" optimality, they may completely bypass "objective" optimality (Einhorn & Hogarth, 1981). Pursuing subjective optimality should not be viewed as a bias since the outcome of reasoning ultimately has to be evaluated according to its utility for the reasoner. Furthermore, whether conclusions are objectively optimal or not is ultimately judged by the reasoner and so intuitive judgements are fundamental to the evaluation process (see also Cohen, 1981; Osherson, 1990). Although not new, these ideas appear to have been neglected by some reasoning theorists. To be sure, not all intuitive judgements will be error free because they may be predicated on inaccurate beliefs, but beliefs in general illuminate the path that reasoning takes. As Newell and Simon (1972) point out, a person's representation of the task is central to understanding his or her behaviour.

Conclusion and Future Directions

The central theme of this chapter was motivated by words from Newell and Simon's (1972) analysis of human problem solving: "[T]o the extent that the behaviour departs from perfect rationality, we gain information about the psychology from the subject, [and] about the nature of the internal mechanisms limiting his performance" (p.55). Newell and Simon's words are fitting because too often it seems theories are imposed on data even when the data repeatedly fail to conform to the theories. When this occurs, it seems wise to step back and reconsider some of the basic assumptions theorists have made and why the data repeatedly fail to confirm those assumptions.

In this chapter, I have attempted to present an alternate hypothesis for interpreting

formal reasoning performance using the inductive-coherence model. Ultimately, the adequacy of this inductive hypothesis will be demonstrated empirically; before this can be done, however, traditional formal assumptions must be flushed out of even the most informal current theories such as pragmatic reasoning theory and mental models. Although proponents of both these theories provide a more inductive and adaptive framework with which to view reasoning, both still implicitly embody traditional formal assumptions about reasoning performance. For example, proponents of both these theories essentially claim that ordinary comprehension processes can interfere with reasoning performance. Proponents of pragmatic reasoning theory argue that participants' will fail to select the correct cards in Wason's task if its context does not invoke the appropriate schema; in other words, if reasoner's ordinary comprehension processes fail to interpret appropriately the context of the task and thus impede the correct schema from being applied. Similarly, mental models theorists maintain that ordinary comprehension processes influence the kind of model created about the premises, which in turn influences the appropriateness of the inference generated.

Interfering factors such as ordinary comprehension processes and working memory limitations can only be presumed to interfere with reasoning if it is assumed at the outset that formal processes underlie reasoning. Under an inductive hypothesis, however, these interfering factors are the basic constraints of an informal, inductive reasoning system designed to utilize background knowledge in order to generate efficient conclusions to ill-defined everyday problems. Under this inductive assumption, errors observed on formal tasks are only "errors" when viewed from a formal perspective. From an inductive perspective, errors are seen as inductive responses to tasks that fail to meet the constraints of a reasoning system adapted to deal with ill-defined problems. Where does this leave us? Just as investigators have rigorously tested the formal assumption, so too must investigators now test the informal, inductive assumption. For example, some researchers have presented accounts of how deductive strategies may arise from an inductive framework (e.g., Sloman, 1998; Sloman & Rips, 1998). Others have proposed how categorizations based on sophisticated and abstract rules may "initially be solved by using

perceptual similarity between items to be categorized and known category members" (Goldstone & Barsalou, 1998, p.244). Although I have offered the inductive-coherence model to understand reasoning performance on formal tasks, I have not described in much detail the kind of architecture that might underlie this model. One inductive framework worth exploring, for example, is a connectionist network because it solves tasks by means of pattern classification, a form of induction.

In the next chapter, I explore how an inductive architecture such as a connectionist architecture can solve Wason's selection task, and how such a solution is organized. As mentioned in Chapter 1, the intent of this section is to show that the appearance of "formal" performance (e.g., solution of Wason's task) need not rely on a formal architecture but can emerge from an inductive architecture.

Chapter 3

Induction as Pattern Classification: Training PDP Networks to Solve Wason's Selection Task

Pattern classification is a form of induction. Although pattern classification is typically associated with concept recognition, it can also be used to describe higher forms of cognition such as reasoning (Bechtel & Abrahamsen, 1991; see also Goldstone & Barsalou, 1998). Proponents of pattern classification accounts of reasoning assume that people make sense of their environment by categorizing objects and events not only to make predictions about their (unseen) characteristics, but also to decide upon actions in light of their categorization. For example, Bechtel and Abrahamsen (1991), reiterating an idea proposed by Margolis (1987), suggest the following pattern classification view of reasoning:

> [T]he recognition of one pattern constitutes an internal cue which, together with the external cues available from outside the system, facilitates yet another recognition. Thus, we work our way through a complex problem by recognizing something, and with the help of that result, recognizing something further. (p.141)

A pattern classification account of reasoning may be explored with Parallel Distributed Processing (PDP) networks, which are trained to solve problems by means of mapping input patterns to output responses. One kind of PDP network is described in the next section.

A PDP Network of Value Units

A PDP network is a system of inter-connected, simple processing units that can be used to classify patterns presented to it. A PDP network is made up of three kinds of processing units. *Input units* encode the stimulus or activity pattern that the network is to classify. *Hidden units* detect features or regularities in the input patterns, which can be used to mediate classification. *Output units* represent the network's response to an input

pattern (i.e., the category to which the pattern is to be assigned) on the basis of features or regularities that have been detected by the hidden units. Processing units communicate by means of weighted connections. Figure 3.1 provides an illustration of a typical PDP network.

In most cases, a processing unit carries out three central functions: First, a processing unit computes the net input or the total signal that it receives from other units. A *net input function* is used to carry out this calculation. After the processing unit determines its net input, it transforms it into an internal level of activity, which typically ranges between 0 and 1. The internal activity level is calculated by means of an *activation function*. Finally, the processing unit determines the signal to be sent to other units. This signal is created by applying an *output function* to the unit's internal activity. The most common output function is the identity function, which suggests that the signal sent out usually equals a unit's internal activity. (The reader is referred to Dawson (1998) for a more complete explication of the different functions). A weighted connection acts as a communication channel between two processing units, and either amplifies or attenuates the signal being sent from one processing unit to another.

In contrast to a conventional computer, a network is not given a step by step procedure for performing a desired task, but is instead *trained* to solve the task on its own. For instance, consider a popular supervised learning procedure called the *generalized delta rule* (Rumelhart, Hinton, & Williams, 1986). To train a network with this rule, one starts with a network (of a pre-specified number of hidden units) that has small, randomly assigned connection weights. The network is then “developed” by presenting it a set of training patterns, each of which is associated with a desired response. To train a network on one of these patterns, the pattern is presented to the network’s input units, and the network generates a response to this stimulus using its existing connection weights. An error value for each output unit is generated by comparing the actual output to the desired output. This error value is then fed backwards through the network, and is used to modify connection weights in such a way that the next time this pattern is presented to the network, the network’s output errors will be smaller. By

repeating this procedure a large number of times for each pattern in the training set, the network's response errors for each pattern can be reduced to near zero. At the end of this procedure, the network will have a very specific pattern of connectivity (in comparison to its random start) and will have learned to perform the desired stimulus/response pairing (if it is possible for such a pairing to be learned).

A number of different versions of the generalized delta rule exist, each designed to train networks whose processors have specific properties. For instance, one form of this rule is applied when the logistic equation is used as an activation function (Rumelhart, Hinton, & Williams, 1986). A slightly different rule is used to train networks of *value units* (Dawson & Schopflocher, 1992). A value unit is a processor that uses a Gaussian equation for the activation function of its processing units.

One advantage of using a value unit network is that its hidden units often exhibit properties which render them interpretable[3] (e.g., Berkeley, Dawson, Medler, Schopflocher & Hornsby, 1995; Dawson, 1998; Dawson, Medler & Berkeley, 1997). As a result, after the network learns to respond to a task, the method by which it maps input patterns to output responses may be examined and assessed for whether it represents both a psychologically interesting and plausible algorithm. Being able to examine the network is important to the current study because understanding how the network solves the task may shed light on questions of why human participants typically fail to respond correctly to the selection task and why they generate alternate responses.

Goals of the Present Study

The first goal of the present study was to train different value unit networks to perform Wason's selection task in qualitatively different ways. For example, one network was trained to respond incorrectly by having it select only the antecedent (i.e., *p* card) in

---

[3]This is often not the case with networks of "integration devices," processing units that use a sigmoid-shaped activation function and have been criticized by researchers as being very hard to interpret (e.g., Berkeley et al., 1995).

response to the task[4]. This response is frequently generated by participants. A second network was trained to respond correctly by having it select both the antecedent and the negation of the consequent (i.e., *not-q* card) in response to the task. The purpose of training these two networks was to show that a solution to Wason's task can be generated by a pattern classification system (i.e., an inductive architecture). To my knowledge this is the first demonstration of a PDP network being trained to solve Wason's task.

The second goal was to interpret the method by which these networks solved the task. Through interpretation, I explore the kind of algorithm that an inductive architecture uses to respond to a deductive task, and why choosing only the antecedent may be an easier response to make than choosing both the antecedent and negation of the consequent. In more succinct terms, the intent was to answer such questions as "Why might human subjects only turn over the antecedent in Wason's selection task?".

Network 1: Selection of the Antecedent

As noted previously, human participants rarely generate the logically correct response to Wason's selection task. One common (but incorrect) response is to select the card representing the antecedent of the rule. The purpose of the first study was to analyze a network trained to generate the antecedent of the rule in an attempt to shed light on the question "Why do human participants make this common mistake?".

Method

PDP version Of Wason's selection task: Network architecture. When human participants are presented with Wason's selection task, they already have a great deal of knowledge about the components of the task. For instance, participants come to the task with knowledge of the connective "if then," of different kinds of numbers (i.e., odd versus

---

[4]I also trained a network to generate the partially incorrect response of "p" and "q" that is also commonly generated by human participants. The results from training this network are not presented here but are supportive of the ideas presented in this chapter. These results may be requested from the author.

even), as well as different letters (i.e., vowels versus consonants). In contrast, PDP networks do not start with this kind of knowledge. Thus, in order to study a PDP network's inductive treatment of the task, the task has to be encoded in some format that the network can process, and the behavioral repertoire of the network at the end of training must be large (or sophisticated) enough to be of interest to psychologists.

The solution to these issues was first to generate a “network friendly” representation of Wason’s selection task, and then to train the network with a sufficiently large training set so that its responses after training were diverse enough to be of interest. A binary code was developed that allowed a representation of both the task’s conditional rule and the four cards using 16 input units. Four inputs units were used to represent the rule. The first two input units reflected the antecedent of the rule, while the last two units reflected the consequent of the rule (see Table 3.1). Four sets of three input units were used to represent the card categories. The first two input units of each set were used to represent the card’s category membership, while the last unit of each set was used to represent its specific instance. Using this encoding scheme, a training set was developed that included eight different conditional rules and eight different symbols (two vowels, two consonants, two even numbers, and two odd numbers) depicting instances of card categories. These instances were crossed with the 24 unique orders generated from assembling 4 cards in all possible combinations. This latter step led to 384 unique orders of card values, which were crossed with each of the eight rules to produce a final training set of 3072 input patterns.

As shown in Figure 3.2, network 1 used three hidden units to learn the task. Following pilot simulations, three was the minimum number of hidden units required for the network to converge, that is, learn the desired mapping between input pattern and output response. Furthermore, the network included four output units, one corresponding to each card. In particular, output unit 1 turned "on" (i.e., was activated to a value of 1) if card 1 matched the antecedent of the rule, but turned “off” if card 1 did not match the antecedent of the rule. Only one output unit was activated in response to an input pattern in the present study because the desired response involved only the selection of the card value affirming the antecedent of the rule.

It is important to reiterate that my interest was not in the process of training the network, but rather in examining the structure of a mature network after training. A mature network responds accurately and reliably to a complete set of training patterns. In this study, reliability of response required that (a) the network be able to identify the presented rule, (b) the network have some representation that assigned each input symbol to a more abstract category (e.g., differentiating between "vowel" and "odd number"), and (c) the network have some representation of the "content" of the presented rule, such that its output would indicate what could be done to test the validity of the rule. Network performance consistent with these three requirements for such a large number of different patterns reflecting Wason's selection task has developed sufficient abilities to be of psychological interest. In short, such a network could generate a psychologically plausible theory of how the selection task could be accomplished by an inductive architecture.

Training. The first network was trained using Dawson and Schopflocher's (1992) elaboration of Rumelhart et al.'s (1986) generalized delta rule, with a learning rate of 0.001 and a momentum of zero. Connection weights and unit biases (i.e., the mean of the Gaussian) were randomly selected from the range of -1.0 to +1.0. The network was trained for 83 epochs (i.e., 83 presentations of the complete training set). The order of pattern presentation was randomized during each epoch so that the network's learning of the task was contingent on the specific input patterns and not on their specific sequence of presentation.

At the end of training, the network generated a "hit" in response to every pattern. A desired response or "hit" consisted of an activation of 0.9 or higher in the output unit corresponding to the "card" affirming the antecedent of the rule along with activations of 0.1 or lower in output units corresponding to cards not matching the antecedent of the rule.

Following training, network performance on the selection task was considered analogous to human performance in so far as the network generated reliable responses to specific input. That is, the network generated the antecedent card in response to the rule in the task, which is consistent with the pattern typically shown by human participants.

Results: Definite Features of Hidden Units

The main goal of this first study was to interpret the structure of a mature network in order to determine the method it used to respond to the task. Accomplishing this goal requires examining each of the network's hidden units for the input features it detects. Once the relationship between hidden units and input features is laid out, is it possible to determine how the hidden units work together in responding to the task. This sequence of analysis is much like putting together a jigsaw puzzle--first one becomes familiar with the individual pieces and, following this, then one is ready to determine how each piece contributes to the whole. The first step of the analysis requires "wire-tapping" the individual hidden units to determine the input features each one discriminates.

"Wire-tapping" is a technique used in recording the activity of hidden units in response to input features (Dawson, 1998; Moorehead, Haig, & Clement, 1989). An analogous technique originated with brain researchers in the 1950s who were interested in recording the behavior of individual neurons (Dawson, 1998). "Wire-tapping" has facilitated interpretation of value unit networks. In particular, following training, the set of input patterns is again presented to the network while the activities that they produce in individual hidden units are recorded. Once this is done, individual hidden units can be examined for the input features they discriminate.

After wiretapping, *jittered density plots* of each hidden unit are constructed as shown in Figure 3.3. A jittered density plot illustrates the *distribution* of activation values produced in a single hidden unit of a mature network following a presentation of the full set of input patterns. A single dot in the plot represents the activation that one input pattern produces in the hidden unit. Hence, one plot illustrates as many dots as there are input patterns. The x-axis on the jittered density plot ranges from 0 to 1 and shows the range of activation values generated by the input pattern set. Dots are also randomly jittered along the y-axis to make them as discernable as possible.

Jittered density plots of value units are frequently highly structured or "banded." The distinct bands suggest the network's hidden units are reliably discriminating specific input features as the network responds to the task. As shown in Figure 3.3, all three of the

network's hidden units exhibit a high degree of banding. Bands render critical information about the specific input features a hidden unit discriminates (Berkeley et al., 1995). Specifically, the bands represent groups of input patterns that share similar features and produce similar activations in a hidden unit. With the aid of descriptive statistics, it is possible to determine the specific features that cluster into each band and how the network uses these sets of features to solve the task. For example, by calculating the Pearson product-moment correlation among the input units of the patterns falling into a band, *definite binary features* can be identified. A definite binary feature indicates a perfectly reliable relationship between input units, for example, a perfectly positive correlation between input unit 6 and 7 indicates that input units 6 and 7 always assume the same value; if input unit 6 has a value of 1.0, then so does input unit 7. In contrast, a perfectly negative correlation between a pair of input units suggests that whenever one input unit is 1.0, the other is 0.0 and vice versa. In the case of network 1, I found that only binary definite features clustered into each band. This finding suggests that hidden units were discriminating relationships among input features and not absolute features[5]. A description of the features detected by each hidden unit is presented in Table 3.2.

Interpretation of definite features: A "matching" algorithm. Network 1 solves the task by means of specialized hidden units; each hidden unit adopts a card and "flips" it over when it matches the antecedent of the rule. For example, imagine the network is presented with a rule and four cards and the 3rd card matches the antecedent of the rule. The network's hidden units check to see if their respective adopted cards match the antecedent of the rule. For example, hidden unit 0 checks card 3, hidden unit 1 checks card 2, and hidden unit 2 checks card 1. Because in this case the 3rd card matches the antecedent of the rule, only hidden unit 0 generates high activity and flips the card over. If

[5]Although not present in the bands shown for network 1, *definite unary features* are also possible to identify by calculating the mean and standard deviation of each input unit collapsed across the entire sub-set of patterns falling in the band. A definite unary feature is a feature shared by all the input patterns in a band. For example, if the mean value for input 5 at band "a" is 1.0 and its standard deviation is 0.0, then this indicates that all the input patterns in band "a" have a value of 1 at input 5.

the matching card had been in 2nd or 1st place, then hidden units 1 and 2, respectively, would have generated high activity and flipped the card over. In contrast, if the matching card had been in 4th place, then none of the hidden units would have generated high activity and this very absence of high activity would have flipped the 4th card over.

The hidden units determine which card matches the antecedent of the rule by exploiting the pattern's specific features. As shown in Table 3.2, each hidden unit detects specific features. For example, all of the patterns in band C of hidden unit 0 are patterns in which input unit 0 has the same activity as input unit 10, and in which input 1 has the same activity as input unit 11. Recall that input units 0 and 1 represent the antecedent of the rule. Similarly, input units 10 and 11 represent the category to which the object of card 3 belongs. When the pattern of activities in input units 0 and 1 is the same as the pattern of activities in input units 10 and 11, hidden unit 0 generates high activity (i.e., band C), and this activity represents the fact that card 3 matches the antecedent of the rule.

Table 3.2 also reveals that an artifact of the match between card 3's category membership and the antecedent of the rule is the mismatch between input unit 0 and input unit 2 as well as the mismatch between input unit 2 and input unit 10. Recall that input unit 2 represents the consequent of the rule. Hence, input units 0 and 2 should always be unequal (i.e., negatively correlated) to each other because none of the rules used in the training set involve equivalent categories. For example, none of the rules are of the form "if vowel, then vowel." By extension, input units 2 and 10 should also mismatch in this case since input unit 10 is equal to input unit 1–one of the units representing the antecedent of the rule.

Bands A and B of hidden unit 0 reflect properties that are true only when band C's properties are false. For example, imagine a pattern in which card 3 does not match the antecedent of the rule. In this case, hidden unit 0 would not show high activity (i.e., band C) because input units 0 and 1 would not be equal to input units 10 and 11, respectively. In this case, however, hidden unit 0-bands A or B would detect the mismatch between card 3 and the antecedent of the rule. For example, band A reflects patterns where input unit 10 is unequal to input unit 0 but equal to input unit 2 (i.e., the first input unit

reflecting the consequent).

Similar accounts follow for the definite features detected by hidden units 1 and 2. The only exception being that the hidden unit 1 detects matches between card 2 (i.e., input units 7 and 8) and the antecedent of the rule, and hidden unit 2 detects matches between card 1 (i.e., input units 4 and 5) and the antecedent of the rule (see Table 3.2).

These analyses indicate that 3 of the output responses are controlled directly by the 3 hidden units of network 1. What about the fourth response? In value unit architectures, it is possible for a zero signal to moderate high activity or a “1" response if both the “bias” of the architecture is equal to zero and the signal being sent is equal to zero. In other words, in a state where none of the three hidden units are activated (i.e., the card matching the antecedent of the rule is not found at card 1, 2 or 3), this state signals a response of “1" in output unit 4.

Discussion. Although each hidden unit identifies a card at a specific location, all three hidden units accomplish this similarly. Notice how the jittered density plots of all three hidden units are alike in their distributions of activation values. This similarity suggests that hidden units respond alike one another in the identification of the respective cards. For example, the first and second band clusters (i.e., bands A and B in all hidden units) consist of patterns that mismatch the antecedent of the rule, while the third band cluster consists of patterns that match the antecedent of the rule. In short, the algorithm by which the network responds to the task proceeds by means of similarly specialized hidden units.

It appears that selecting the card affirming the antecedent of the rule is accomplished fairly easily, that is, the network seems to just be matching the cards to the antecedent of the rule. The network’s “matching” algorithm is somewhat similar to the algorithm that human participants have been shown to use. For example, Manktelow and Evans (1979; and Evans, 1989) provide evidence to suggest that participants’ responses to the selection task can be explained by a model whereby selected responses simply match the antecedent and consequent of the rule. The algorithm may also be viewed as a “similarity” algorithm where the network learns to select the response that shares the most features with the antecedent of the rule. A number of investigators (e.g., Goldstone & Barsalou, 1998;

Sloman & Rips, 1998; Tversky & Kahneman, 1974) support the role of similarity as a gateway to higher-level cognitive capacities.

Network 2: Selection of Antecedent and Negation of Consequent

Method

Network architecture. The second network that was trained used the same input unit encoding and training patterns as the first network. Also, output units in network 2 were organized as in network 1. Two output units, however, instead on of one were designed to turn "on" in response to every input pattern; one output unit to show the card matching the antecedent of the rule and another output unit to show the card negating the consequent of the rule. Finally, as shown in Figure 3.4, network 2 consisted of eight hidden units instead of three because this was the minimum number required for the network to converge.

Training. Network 2 was trained following the same procedure used with network 1. As with network 1, prior to training, the network's connection weights were randomly set to values between -1.0 to +1.0, while its unit biases were set to 0.1. The learning rate was 0.001 and no momentum was used. At the end of training, the network generated a "hit" in response to every pattern. A desired response or "hit" consisted of an activation of 0.9 or higher in the output units corresponding to the cards affirming the antecedent and negating the consequent of the rule, and activations of 0.1 or lower in output units corresponding to the cards negating the antecedent and affirming the consequent. The network learned to generate the desired response to all patterns after 414 epochs of training.

Results

Interpretation of definite features: A joint effort in matching. The analysis of network 2 proceeded similarly to network 1. As shown in Figures 3.5 and 3.6, all 8 jittered density plots exhibited a high degree of banding suggesting that the network's solution to the task

involved the detection of definite features. What is striking about network 2's jittered density plots is that *pairs of hidden units* have similar patterns. As shown in Tables 3.3 and 3.4, hidden units 0 and 6, hidden units 1 and 4, hidden units 2 and 5, and hidden units 3 and 7 all bear strong correlations with each other when a specific card is selected. For example, hidden units 3 and 7 share a correlation of -.994 when card 1 is a desired response. Likewise, hidden units 2 and 5 share a perfectly positive correlation when card 3 is a desired response. These strong correlations disappear when the corresponding card does not represent a desired response.

The relation between pairs of hidden units is important for generating correct responses. For example, consider the pair of hidden units 2 and 5. As shown in Table 3.7, hidden unit 2 shows high activity (i.e., band E) when input unit 11 has the same activity as input unit 1 but not input unit 3, and input units 0 and 2 are unequal to each other as are input units 1 and 3. Recall that input units 0 and 1 represent the antecedent of the rule, while input units 2 and 3 represent the consequent of the rule. These definite features indicate that patterns whose rules are 0011, 1100, 0110, or 1001 and show card 3 to affirm the antecedent or negate the consequent of the rule are detected by hidden unit 2-band E. In addition, hidden unit 2-bands A and B reflect properties that are true only when band E's properties are false. For example, bands A and B detect patterns that reflect either a mismatch between card 3 and the antecedent of the rule or a match between card 3 and the consequent of the rule.

Hidden unit 5 also shows high activity (i.e., band C) when input units 0 and 2 are unequal to each other and input units 1 and 3 are unequal to each other. These definite features suggest that patterns reflecting the rules 0011, 1100, 0110, and 1001 (the same ones detected by hidden unit 2-band E) are detected by hidden unit 5-band C. Although not apparent from the definite features, hidden unit 5 also shows high activity (i.e., band C) whenever card 3 either affirms or negates both the antecedent or the consequent of the rule. In this respect, hidden unit 5-band C works primarily at detecting specific rules and picking out card 3 whenever it is convenient to do so.

Furthermore, hidden unit 2 shows high activity (i.e., band D) when input units 10 and

11 are unequal, when input units 0 and 2 are unequal, and when input units 1 and 3 are equal to each other. These features suggests that patterns whose rules are 0111 and 1000 and show card 3 to affirm the antecedent of the rule or negate the consequent of the rule are detected by hidden unit 2-band D. Hidden unit 5 also shows activity (i.e., band B) in response to patterns reflecting the rules 0111 and 1000, and show card 3 to either affirm or negate both the antecedent or the consequent of the rule. In contrast, hidden unit 2-band C reflect properties that are true only when hidden unit 2-band D's properties are false; that is, when card 3 fails to match the antecedent or negate the consequent for rules 0111 and 1000.

Finally, hidden unit 2 also shows activity (i.e., band C) when input units 10 and 11 equal each other, input units 1 and 3 equal each other, but input units 0 and 2 are unequal. These features indicate that patterns whose rules are 0010 and 1101 and show card 3 to affirm the antecedent of the rule or negate the consequent are detected by band C. Likewise, hidden unit 5 shows activity (i.e., band A) in response to patterns whose rules are 0010 and 1101 and show card 3 to either match or mismatch both the antecedent or consequent of the rule. In contrast, hidden unit 2-band D also reflects properties that are true only when hidden unit 2-band C's properties are false; that is, when card 3 fails to match the antecedent or negate the consequent for rules 0010 and 1101. Similar accounts may be made for the remaining pairs of hidden units (see Tables 3.5, 3.6, and 3.8).

Discussion. These analyses indicate that output responses are controlled directly by pairs of hidden units. In particular, hidden units 0 and 6 together activate card 2 when it is appropriate to do so, hidden units 1 and 4 activate card 4, hidden units 2 and 5 activate card 3, and hidden units 3 and 7 activate card 1. In addition, as with network 1, the jittered density plots for each of these sets of hidden units have similar distributions, suggesting that these pairs of units are responding similarly to the set of training patterns.

Although network 2 employs a more sophisticated algorithm than network 1, hidden units in both networks discriminate all rules and card values at specific locations. As with network 1, furthermore, network 2's algorithm may be characterized as a matching algorithm. For example, pairs of hidden units jointly discriminate among the training

pattern's "cards" to find the one card that matches the antecedent of the rule and the other that negates the consequent of the rule.

General Discussion

The goals of this study were to (a) train different value unit networks to perform Wason's selection task in qualitatively different ways (i.e., one correct but incomplete response and another completely correct response), and (b) interpret the method by which the networks solved or responded to the task.

The two simulations demonstrated that an inductive architecture could, in practice, solve different representations of Wason's selection task. For example, network 1 was trained to select only the card affirming the antecedent of the rule, while network 2 was trained to select both the card affirming the antecedent and the card negating the consequent of the rule. Results from training both networks suggest that selecting both cards is a more difficult response to generate than selecting only one card. Difficulty here is manifested by the number of hidden units required by the network to solve the task (i.e., network 1 required three while network 2 required eight hidden units to converge). In addition, network 1 required 83 epochs to learn the desired input-output mapping while network 2 required 414 epochs to learn the mapping. One immediate critique is that network 2 was bound to require more hidden units and a greater number of epochs to converge since its task is more complex, that is, network 2 must select two cards instead of one. If this were the case, however, one could have expected network 2 to require only six hidden units and 166 epochs to converge (i.e., double the requirement of network 1). In fact, network 2 required more than this. It required more than twice the number of hidden units and almost 5 times the number of epochs to reach convergence. In sum, the task given to network 2 appears more difficult than the task given to network 1. Hence, one reason why many participants may select only the "p" card instead of both the "p" and "not-q" is because this is an easier response to generate. Learning to generate only the "p" card requires fewer hidden units and fewer epochs than learning to generate both the "p" and "not-q" cards.

My second goal was to understand how an inductive architecture could, in practice, solve different representations of Wason's selection task. The results from wire-tapping the hidden units of network 1 and 2 indicated that both networks used similar algorithms to generate desired responses. Both networks employed an algorithm that generated the desired output via specialized hidden units discriminating all rules but only a single card location. For instance, each of network 1's hidden units discriminated all rules but discriminated card values at a single location. In contrast, network 2's hidden units paired up and each set discriminated all rules but card values at a single location.

The algorithm generated by network 2 suggests that a logical response may be generated without the need for logical rules. Network 2's highly specialized algorithm wherein pairs of hidden units discriminate all rules but card values at a single location indicates an absence of general (formal) rules and the importance of specific instances (i.e., each of the task's rules and card values) for solving the task. The suggestion that specific instances are important for solving a deductive-style problem such as Wason's selection task complements inductive proposals of reasoning; in particular, inductive proposals that stress the role of familiarity or context-specific reasoning algorithms (e.g., Cheng & Holyoak, 1985). Viewing reasoning as pattern classification also supports theoretical claims that prior experience, either by affording familiarity or conceptual frames in which information is categorized and easily interpreted, is critical to reasoning because it alerts reasoners to important variables for solving tasks (Bisanz et al., 1994; Cosmides & Tooby, 1996; Holland et al., 1986). To my knowledge, this is the first demonstration of an inductive architecture trained to perform as if it had logical rules. In sum, both of the goals set out in this study were accomplished.

Finally, although these results suggest that the Wason selection task may be better conceptualized as an inductive task that elicits inductive reasoning, I do not presume here that formal rules do not play an active role in reasoning. There are many examples to suggest that they do as when we confidently deduce a conclusion from a set of premises. It is possible, however, that deductive strategies may arise from an inductive framework (see for e.g., Sloman, 1998; Sloman & Rips, 1998). For example, Goldstone and Barsalou

(1998) propose that categorizations that are based on sophisticated and abstract rules may "initially be solved by using perceptual similarity between items to be categorized and known category members" (p.244). The point stressed here is that in general the architecture of reasoning may be better conceptualized as inductive given the nature of the problems that need to be solved in our everyday environment. By conceptualizing all reasoning as originating from inductive origins, we may start to ask different questions and pursue different avenues toward a fuller understanding of reasoning.

The results presented in this chapter outline the kind of architecture that might underlie the inductive-coherence model. In addition, although these results clearly support the hypothesis that human reasoning may involve an inductive architecture, yet another approach to testing the inductive hypothesis would be a demonstration that human behaviour is influenced when inductive variables are varied within a formal reasoning task. In the next chapter, inductive variables are manipulated within a thematic version of Wason's task in order to show that formal performance is controlled via inductive processes.

Chapter 4

The Effects of Manipulating Inductive Variables Within a Thematic Version of Wason's Selection Task

Framing Wason's task in a *thematic* context (instead of its original abstract context) has shown to facilitate participants' logical performance on the task; that is, yield a *thematic effect* (Cox & Griggs, 1982; Evans et al., 1993). This effect, however, has not been consistently attained, prompting a number of explanations about when a thematic version will lead to logical performance. For example, Cheng and Holyoak (1985, 1989) in their pragmatic reasoning theory suggest that versions of the task involving a permission context, along with a participant's goals, elicit a permission schema that facilitates logical performance.

Other investigators have proposed similar theories but with varying details. For example, Cosmides (1989; also Gigerenzer & Hug, 1994) argues that a permission context is not sufficient for inducing the thematic effect, but instead a *social contract* context that evokes a "detection of cheaters" algorithm, is necessary to induce the effect. Other more comprehensive theories of human reasoning have also provided accounts of thematic effects (see Johnson-Laird & Byrne, 1991), although these have not been as influential as pragmatic reasoning theory and social contract theory. Implicit in pragmatic reasoning theory as well as social contract theory is that performance on Wason's selection task is accomplished largely via inductive processes since proponents of both theories claim that the context of the task is fundamental to evoking the appropriate schema (or algorithm in Cosmides' theory). According to these theorists, participants perform poorly on abstract versions of the task because the permission schema (or detection of cheaters algorithm) is not induced by the task's content (Cheng & Holyoak, 1986, 1989; Cosmides, 1989).

Although supporters of both these theories suggest that context is fundamental to task performance, they do not mention that familiarity with the task might also be important. In fact, these theorists differentiate their accounts (i.e., pragmatic reasoning theory and social

contract theory) from the *memory-cuing hypothesis* (Cox & Griggs, 1982) of thematic effects. The memory-cuing hypothesis is a "familiarity account" of thematic effects, in which facilitated logical performance is attributed to contexts that are able to evoke participants' memories of desired responses. For example, supporters of this hypothesis explain participants' logical performance on the "drinking problem" (Griggs & Cox, 1982) by suggesting that the context of the task elicits participants' memory of similar situations and, as a result, the correct responses. In contrast, pragmatic reasoning and social contract theorists suggest that improved logical performance on the drinking problem results from its specific contextual structure (e.g., a permission situation), which evokes a domain-specific algorithm.

Despite efforts to differentiate pragmatic reasoning theory and social contract theory from the memory-cuing hypothesis, it is unclear whether these accounts are truly distinct. This is because the thematic tasks used to test pragmatic reasoning theory and social contract theory also function as tests of the memory-cuing hypothesis. Specifically, the *unfamiliar* tasks that have been used to test social contract theory and pragmatic reasoning theory also test the memory-cuing hypothesis since the task's unfamiliar context is not very unfamiliar and so could be cuing participants' memories. For example, one of Cosmides' (1989) unfamiliar tasks involved a story about a native tribe called the Kaluame who lived under a strict law in which men could not eat a rare aphrodisiac called "cassava root" unless they had a tattoo on their face. In her task, participants were given the following conditional rule along with four cards representing the activities/attributes of four men:

Rule: If a man eats cassava root, then he has a tattoo on his face.

Cards: (1) Eats cassava root (2) Eats molo nuts (3) Tattoo on face (4) No tattoo on face

After reading the rule and cards, participants were instructed to select those cards that showed potential violations of the rule. Close to 75 percent of participants selected the logically correct cards: *Eats cassava root* (i.e., p) and *No tattoo on face* (i.e., not-q)

(Cosmides, 1989). Although Cosmides interpreted this finding as evidence for her "detection of cheaters" algorithm, this is not the only interpretation of the data. This context about the Kaluame tribe, while being unfamiliar in surface details, is similar to common scenarios in which people must first do something before they can do something else. For example, people must first be diagnosed with a certain illness before they can obtain certain prescription drugs. It is possible that the participants in Cosmides' study chose the logically correct cards not because the story evoked a domain-specific algorithm but instead because the story cued participants' memories about other such everyday scenarios where one must satisfy a constraint before pursuing further action. In short, although pragmatic reasoning and social contract theorists maintain that specific contextual factors evoke specialized inductive algorithms, it is unclear whether these factors truly evoke algorithms or simply evoke available memories of familiar situations.

One way to test the largely inductive schema/algorithm hypothesis against the familiarity hypothesis is to manipulate specific variables that should affect inductive reasoning without affecting memory recall. Please note that testing the inductive hypothesis against the familiarity hypothesis does not follow from the inductive-coherence model of reasoning because, in this model, familiarity (or general knowledge) with a task is considered practically inseparable from reasoning about a task. This is because a reasoner's knowledge is considered a fundamental variable to the operation of inductive reasoning processes (Holland et al., 1986). However, as mentioned previously, familiarity and inductive explanations of performance are considered distinct according to some theorists and therefore this distinction is empirically tested.

One variable that has been shown to influence analogical reasoning, a form of induction, is *conceptual distance* (i.e., similarity) (Tourangeau & Sternberg, 1981, 1982). For example, Tourangeau and Sternberg (1981, 1982) suggest that participants evaluate a metaphor's aptness by judging the similarity between the terms of the metaphor within a conceptual pattern space. In support of similarity judgements, Goldstone and Barsalou (1998) have also proposed that categorizations that are based on sophisticated and abstract rules may "initially be solved by using perceptual similarity between items to be

categorized and known category members" (p.244; see also Sloman & Rips, 1998). Following both Tourangeau and Sternberg (1981, 1982) as well as Goldstone and Barsalou (1998), I hypothesized that if participants are solving Wason's selection task inductively, they may do so by invoking a conceptual space. If participants do invoke a conceptual space, I further hypothesized that this space should be more easily invoked when the task is framed in a meaningful, thematic context. A thematic context should facilitate creation of the space because it would cue participants' background knowledge or conceptual frame. In contrast, I hypothesized that this space would not be as easily invoked when the task is framed in an abstract context because the absence of context would be unable to cue (as easily) a participant's background knowledge.

I further assumed that the pattern space invoked to a thematic version should facilitate performance when the cards representing the rule's antecedent (i.e., p and not-p) are judged dissimilar to each other as are the cards representing the rule's consequent (i.e., q and not-q). The reason for assuming that dissimilarity between the rule's antecedent cards (and consequent cards) should facilitate performance stems from the idea that dissimilarity might motivate the creation of a larger pattern space because the participant has to invoke a space that encompasses two opposite antecedent cards. Moreover, a large pattern space might enable participants to consider a greater number of interpretations of the task. For example, given the task to test the following rule:

Rule: If the person goes to the movies, then he/she must be sad.

Cards: (1) Goes to movie (2) Stays home (3) Is sad (4) Is euphoric

Participants might find it easier to test this rule because the cards representing the antecedent are fairly dissimilar (according to the dimension "places to see a show on a Saturday afternoon" versus another dimension such as "places to be at on a Saturday afternoon") as are the cards representing the consequent (according to the dimension mood). In this case, participants might find it easier to appreciate that the cards "Goes to movie" and "Is euphoric" can falsify the rule. An appreciation of the counter-examples or

alternate interpretations to the task's rule has been cited as necessary to select the instances that falsify the rule (Johnson-Laird & Byrne, 1991). In contrast, given the same task but now with similar cards representing the antecedent as well as the consequent of the rule:

Rule: If the person goes to the movies, then he/she must be sad.

Cards: (1) Goes to movie (2) Goes to theater (3) Is sad (4) Is gloomy

Participants in this case might find it harder to select the cards "Goes to movie" and "Is gloomy" to falsify the rule because the choices are almost identical to each other. For example, a participant might think that "sad" and "gloomy" are really the same thing so the choice becomes trivial. Of course, it is important to note that the facilitative effect of a larger pattern space should manifest itself in the frequency with which participants' select the "not-q" card since this is the card that participants have trouble with; not the "p" card.

In order to test conceptual distance without testing familiarity, the context of the task needs to be thematic but not familiar so that it does not cue participants' memories of the context or memories of similar contexts. This way, if conceptual distance does have an effect on performance it cannot be attributed to memory recall.

In addition, a survey of the literature reveals a confound in tests of abstract versions versus thematic versions of Wason's task that may be disentangled by testing the inductive versus familiarity hypothesis (Cheng & Holyoak, 1985, 1989; Cosmides, 1989; Griggs, 1983; Griggs & Cox, 1982; Evans et al., 1993; Johnson-Laird & Byrne, 1991; Wason, 1966). The confound is that studies of abstract versions of Wason's task typically employ *category* conditional rules with *instance-based* cards whereas studies of thematic versions of Wason's task employ *instance-based* conditional rules with *instance-based* cards. For example, researchers who have used abstract versions normally employ category rules such as "if there is a vowel on one side then there is an even number on the other side" along with cards showing instances of the categories represented in the rule (i.e., A, K, 4, and 5) (Evans et al., 1993; Wason, 1966; for a study that does not employ category rules

see Evans & Lynch, 1973). In contrast, researchers who have used thematic versions consistently employ instance-based rules such as “If a man eats cassava root, the he has a tattoo on his face” along with cards closely *matching* the instances shown in the rule (i.e., eats cassava root, eats molo nuts, tattoo on face, no tattoo on face. Given this *abstraction* confound, it is difficult to state conclusively that facilitated performance on thematic versions originates from the thematic context of the task. It is possible that facilitated performance on thematic versions stems from the ease with which participants map the instance-based cards to the instance-based terms in the rule. Conversely, poor performance on abstract versions could stem from the additional mental step of mapping the instance-based cards to the category terms in the rule.

Associated with this confound is another confound. Employing category rules forces one to incorporate cards representing instances from *four* distinct categories. For example, in the category rule, if there is a vowel on one side then there is an even number on the other side, along with the instance-based cards, A, K, 4, and 5, notice that these instances represent four unique categories: vowel, consonant, even number, and odd number. Given a category rule, one must have four categories represented in the cards or else the task does not make much sense. For example, if one had two categories, say two instances of a vowel and two instances of an even number, then one essentially has two “p” cards and two “q” cards without instances of a “not-p” and a “not-q” card. This situation would not make sense in the context of the task. In contrast, instance-based rules are consistently tested with cards representing instances from only *two* distinct categories. For example, in the instance-based rule, if a man eats cassava root, the he has a tattoo on his face, along with the instance-based cards eats cassava root, eats molo nuts, tattoo on face, and no tattoo on face, notice that these instances represent essentially two unique categories: dinner and facial identification. Given this additional confound, it is again difficult to state conclusively that facilitated performance on thematic versions originates from the thematic context of the task. It is possible that facilitated performance on thematic versions stems from the different number of categories represented in the cards. For example, participants might find it easier to think of the falsifying cards when they only have to consider two

categories than when they have to consider four. The relation between these two confounds is shown in Figure 4.1. I attempt to clarify these confounds in Study 1.

Whereas *conceptual distance* was identified earlier as an inductive variable, the *abstraction* and *number of (card) categories* associated with the conditional rule are also variables that test the hypothesis that thematic and abstract versions of Wason's task are solved via inductive processes. The abstraction and number of (card) categories associated with the conditional rule should only have effects on participants' performance if Wason's task is solved inductively. Under formal theories of human reasoning (especially syntactic theories), only the syntax or structure of the conditional rule should affect reasoning (e.g., Braine & O'Brien, 1991; Rips, 1994, 1995). In these theories, whether the conditional employs category or instance-based terms or whether the cards represent two or four categories should not alter performance because these variables are contextual and not syntactic. Moreover, effects stemming from the abstraction and number of (card) categories associated with the conditional rule cannot be attributed to memory recall because these variables should not cue recall of specific situations.

Study 1

In this Study, *conceptual distance* was manipulated within a thematic version of Wason's task. My first hypothesis was that an increased conceptual distance both between the cards representing the antecedent and between the cards representing the consequent should facilitate participants' selection of the cards that falsify the conditional rule. Furthermore, both the *abstraction* of the conditional rule (i.e., category versus instance-based) and the *number of categories* represented in the cards were manipulated within a thematic and abstract version of Wason's task. My second hypothesis was that category rules may hinder participants' selection of the cards that falsify the conditional rule because in this condition participants need to map the rule's category terms to the instances represented in the cards. This extra mental step may provoke participants to make mapping errors. Conversely, I hypothesized that instance-based rules may facilitate participants' selection of the cards that falsify the rule because in this condition the cards

are more easily mapped against the terms in the rule. My third hypothesis was that participants would be more likely to select the cards that falsify the rule when card instances are generated from only two categories instead of four. Because I expected two categories to occupy less working memory capacity, participants might be more likely to consider all cards and, therefore, to choose the correct cards with which to falsify the rule. In contrast, four categories might occupy more working memory capacity, leading participants to only partially consider the cards that falsify the rule.

Method

Participants. Fifty-seven students from introductory psychology courses at the University of Alberta served as participants. All participants received class credit for collaborating, and none had training in logic.

Materials. In order to test the variables of conceptual distance, abstraction and number of (card) categories, a *Batman thematic task*[6] was developed and is shown in Figure 4.2 (both the task and a sample of 4 rules in each variable condition are shown in Appendix 1). The Batman context seemed appropriate because its theme is meaningful but likely unfamiliar to many participants; that is, few participants would have had direct or even indirect or *analogous* experience with this problem solving scenario. Direct or indirect experience with the context of the task was purposefully avoided so that memory-recall explanations could be circumvented if the manipulations had any effect on participants' performance. As shown in Figure 4.2, the task involves testing a series of conditional rules that the Joker has created involving the contents and name plates of rooms at a museum. The evidence used to test the rules comes in the form of rooms, each with either a closed or open door. For example, given a rule of the form, *If the name plate refers to a sow, then there is a bluejay inside the room*, and 2 closed rooms with *sow* and *goat* on the name plates, and 2 open rooms with *eagle* and *bluejay* inside, the participant must select which rooms to open and/or close in order to test the rule. Before describing the

---

[6]I thank Professor Gay Bisanz, a member of my supervisory committee, for suggesting the Batman context.

procedure participants followed on this task, a description of the variables and how they were manipulated is presented.

Design. First, conceptual distance was manipulated according to the method described by Tourangeau and Sternberg (1981, 1982). Conceptual distance was broken down into (a) the distance between the antecedent and consequent in the conditional rule (e.g., if *antecedent* then *consequent*), and (b) the distance both between the card instances representing the antecedent and the card instances representing the consequent (see Figure 4.3 for an illustration). Similar to the terminology used in Tourangeau and Sternberg (1981, 1982), *between-term distance* describes the distance between the antecedent and consequent in the conditional rule and *within-term distance* describes the distance between the card instances representing the antecedent and the distance between the card instances representing the consequent.

Toglia's and Battig's (1978) word norms were used to generate *high* and *low* within-term distances. The nouns chosen differed along two dimensions: familiarity and pleasantness. Distance between the terms was calculated using the following equation:

$$d^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2 \quad (1)$$

where x and y refer to the familiarity and pleasantness dimensions, respectively. A high within-term distance was defined as a distance (d) of 1.2 or more, while a low within-term distance was defined as a d of 0.5 or less. Between-term distance had to be manipulated differently because category norms were unavailable. As a result, between-term distance was manipulated by invoking two higher-level categories: Living and non-living things. As shown in Figure 4.4, within each of these higher-level categories, I set up 4 categories that were considered to be *low* in between-term distance among one another (e.g., Living things included land animals, birds, aquatic animals, costumed professions). A *high* between-term distance was achieved by taking categories from each of the two higher-level categories (see Figure 4.4 for an example). Given the creation of these higher-level categories, the instances from Toglia and Battig's word norms were chosen so that they

would correspond to the higher-level categories created. Conceptual distance was a within-subjects factor.

Second, the *abstraction* and *number of (card) categories* associated with the conditional rule were manipulated in three conditions instead of four because, as shown in Figure 4.1, a category conditional rule cannot be associated with cards representing only two categories. The *first condition* involved a category-based rule with cards reflecting instances from 4 distinct categories. For example, participants in condition 1 read "category" rules of the form, *If the name plate refers to a land animal, then there is a bird inside the room*, and tested this rule by selecting "doors" representing instances from 4 categories, *dog* (i.e., category 1-land animal), *trout* (i.e., category 2-aquatic animal), *bluejay* (i.e., category 3-bird), and *magician* (category 4-costumed profession). The *second condition* involved an instance-based rule with cards reflecting instances from 4 distinct categories. Participants in the second condition read "instance" rules of the form, *If the name plate refers to a dog, then there is a bluejay inside the room*, and tested this rule by selecting doors representing instances from the same 4 categories as in condition 1, *dog*, *trout*, *bluejay*, and *magician*. The *third condition* involved an instance-based rule and cards representing instances from 2 categories. Participants in the third condition read "instance" rules of the form, *If the name plate refers to a dog, then there is a bluejay inside the room*, and tested this rule by selecting doors representing instances from 2 categories, *dog* and *goat* (i.e., category 1-land animal), and *bluejay* and *robin* (category 2-bird). More examples of the three conditions are shown in Appendix 1.

Notice that the order of conditions reflects an ordering of information load. For example, condition 1 reflects the greatest information load on participants because they must contend with a category rule and cards representing instances from four categories. In contrast, condition 3 reflects the lowest information load because participants face an instance-based rule along with cards representing instances from only two categories. Finally, condition 2 represents a mid-sized information load compared with conditions 1 and 3 because it involves an instance-based rule along with cards representing instances from four categories. Abstraction and number of (card) categories were between-subjects

factors.

Procedure. Participants were tested in groups of 10 to 15. Booklets were prepared that included both the Batman task and a standard version of Wason's selection task under each of three *abstraction* and *number of (card) categories* conditions (see Appendix 1). Within each condition, booklets contained 30 rules (along with cards). The Batman task consisted of 6 trials of each of the four combinations of within- and between-term conceptual distance, leading to a total of 24 rules and cards. Two orders of presentation were constructed. The first order was constructed by randomly selecting 24 rules that met both the distance (i.e., high within-term distance of 1.2 or higher and a low within-term distance of 0.5 or less) and valence combination constraints (i.e., 6 trials of each of high between-term and within-term distance (HH), HL, LH, and LL between-term and within-term distance, respectively). The second order was constructed by flipping the first order so that the first rule presented under order 1 was the last rule (i.e., rule #24) presented under order 2.

After completing the Batman task, participants in each condition completed six trials of Wason's task. As with the Batman task, 2 orders of Wason's task were constructed. I presented Wason's task consistently after the Batman task because I wanted to avoid contaminating responses on the Batman task by cuing participants' memory of the Wason's task (many students are exposed to Wason's task in their first year introductory psychology courses). I also did not counterbalance Wason's task with the Batman task because I would not have had a sufficient number of participants in each variable condition for subsequent analyses. Given its abstract nature, conceptual distance was not manipulated in Wason's task, however, abstraction and number of (card) categories associated with the conditional rule were manipulated (see Appendix 1 for examples). For example, in condition 1, Wason's task involved a "category" rule (i.e., *If there is a vowel on one side, then there is an even number on the other side*) along with cards representing instances from 4 categories (i.e., *A*, *K*, *4*, and *7*). Wason's task was included as the "abstract" baseline measure from which to assess thematic effects on the Batman Task.

The investigator instructed participants orally at the beginning of a test session that

they would be completing a booklet containing two tasks. The investigator went on to describe that the first task (i.e., Batman task) included 24 questions, while the second task (i.e., Wason's selection task) included six questions. Participants were asked to read both the instructions and the story carefully and to ask about any unclear instructions. Participants were also asked not to look at the second task until they had finished the first task. All instructions were also reproduced in text form at the beginning of the booklet. After completing the task, participants were debriefed aloud and given a debriefing form to take home. Participants completed the task in groups of 15 and finished in approximately 35 minutes.

Results

Overall frequency of responses. In general, participants's responses to the Batman task were incorrect. Across the three conditions, only 4 to 11 percent of participants chose the correct doors, representing "p" and "not-q." Furthermore, across the three conditions, the three most common door selections out of all possible selections to each of the rules included the door representing the "p" card (19% to 28%), the doors representing the "p" and "q" cards (32% to 44%), and all the doors (12% to 16%) ($\chi 2>42$, df>5; $p=0.00$). Participants' performance on Wason's task mirrored performance on the Batman task since the three most common responses on Wason' task were the "p" card (18% to 25%), the "p" and "q" cards (42% to 47%), and all cards (11% to 12%). Note, however, that selection of the "p" and "not-q" cards was slightly higher on Wason's task (12% to 18%) than on the Batman task.

Conceptual distance. Participants responded consistently to all 24 rules irrespective of between-term distance and within-term distance. An informal inspection of responses revealed that 70 percent of participants selected the same door(s) to 75 percent of rules. Moreover, a Friedman two-way analysis of variance was used to determine if conceptual distance produced any significant effect on participants' responses. This analysis is a nonparametric analog of a repeated measures analysis of variance (Marascuilo & Serlin, 1988). Although this analysis assumes the dependent variable is at least on an ordinal

scale, I decided to use it despite knowing that door selections did not represent an ordinal scale; I did not intend to make any inferences regarding participants' door selections, but merely wanted to determine if participants' responses were consistent. Analysis of participants' responses with the Friedman test revealed that participants within each of the three conditions associated with rule abstraction and number of (card) categories reacted consistently to all rules (condition 1, X_R^2=15.26, df=23, *p*=.89; condition 2, X_R^2=3.30, df=23, *p*=1; and condition 3, X_R^2=3.30, df=23, *p*=1). Participants' responses to all six rules in Wason's task were also consistent (conditions 1, 2, and 3, X_R^2<.4737, *p*>.9931). These results do not support the hypothesis that increases in conceptual distance--between the antecedent and consequent of the rule, and/or between the cards representing the antecedent or consequent of the rule--facilitate logical performance.

Rule abstraction and number of (card) categories. Rule abstraction and number of card categories had no effect on the *type* of doors or cards selected in response to either the Batman task or Wason's task, respectively (i.e., participants within all three conditions tended to choose one of either "p," "p and q," or "all cards" consistently for a majority of rules). In this respect, the hypothesis that performance should be facilitated as the abstraction between rule and cards decreased, and as the number of categories represented in the cards decreased, was not supported.

Although not included in my initial hypotheses, there were differences among the groups in the *number* of doors selected in response to Wason's task. Although this pattern was also found in the Batman task, it was not statistically significant (all $\chi 2$<8, df=4, *p*>.11 for 21 out of 24 rules). A Pearson Chi-square analysis revealed that participants in conditions 1, 2, and 3 differed in the number of doors they selected to test the conditional rule in Wason's task. For example, 92 percent of participants in condition 1 chose at least two or more doors to test the conditional rule, whereas only 78 percent of participants in condition 2 and only 58 percent of participants in condition 3 chose this number of doors ($\chi 2$>10; df=4; *p*<.032 for five out of the six trials; one trial approached significance with a $\chi 2$=9 and *p*=.0663). Alternatively, participants in condition 3 tended to choose fewer doors to test Wason's rule than participants in condition 2, who in turn tended to choose fewer

doors than participants in condition 1. Table 4.1 shows the breakdown of responses by rule abstraction, number of (card) categories, and task. I discuss this unexpected effect in the next section.

Discussion and Conclusion

Conceptual distance. Manipulating conceptual distance did not influence participants' responses on the Batman task. At first glance, this null effect neither supports the formal hypothesis nor the inductive hypothesis. It does not support the formal/logical hypothesis because participants failed to ignore the context and select the correct doors. It also does not support the inductive hypothesis because the meaningfulness of the theme did not facilitate participants' performance. However, this null effect may underscore the importance of familiarity or prior knowledge in reasoning. For example, that manipulations of conceptual distance did not facilitate participants' logical responses may suggest that conceptual distance, within an unusual and unfamiliar context that places high demands on working memory, does not influence responding. It is possible that having background knowledge about a task is pre-requisite for an inductive variable such as conceptual distance to influence performance. This prediction needs further empirical support. It is also possible that the words obtained from Toglia and Battig's (1978) norms did not permit a powerful enough manipulation of distance. That is, these words may have been too close together within a conceptual space to be meaningfully separated by participants. As a result, the manipulation of conceptual distance may not have been a fair measure of the inductive hypothesis.

Rule abstraction and number of (card) categories. As with conceptual distance, rule abstraction together with number of (card) categories did not affect participants' responses to the Batman task. In particular, these variables did not affect the kind of doors or cards chosen to test the conditional rules in the Batman task or Wason' task, respectively. However, an unexpected effect emerged from the manipulation of rule abstraction and number of (card) categories. These variables did affect the number of cards chosen by participants to test the conditional rules in Wason's task. In particular, as

information load increased (as measured by the mismatch between rule and cards and the number of categories represented in the cards) so did the number of cards selected to test the conditional rule. This result may be explained from an inductive perspective. It is possible that as information sources increase, participants are motivated to use more of the available evidence to test a hypothesis because (a) participants are made aware of more information, and (b) participants' confidence in any one piece of evidence decreases given the increase in the information sources to consider (Elio, 1997; Sloman, 1994; Stevenson & Over, 1995). As information increases and confidence decreases, participants may adopt a "safe" strategy where more evidence is chosen with which to test the hypothesis.

One reason why rule abstraction and number of (card) categories did not influence participants' performance on the Batman task might lie in this task's highly unusual and unfamiliar context. Given the unusual character of the contextual information and the high number of unrelated categories presented in the rules, participants in all conditions might have become overwhelmed with information and slipped into a single mode of responding (i.e., heuristic responding) as described in the inductive-coherence model of reasoning. In comparison, Wason's task might have appeared much easier (given that there is less information to attend to at one time) and as a result might have invoked a different approach from participants; an approach that allowed the variables of rule abstraction and number of (card) categories to have an effect.

It is important, however, to focus on the significance of these effects. First, a lack of thematic effects on the Batman task provides further support that not all thematic tasks produce facilitation. Second, the absence of effects from manipulations of conceptual distance and rule abstraction and number of (card) categories on the Batman task suggests that more study needs to be devoted to the interaction between context and familiarity. Finally, an effect from the manipulation of rule abstraction and number of (card) categories on Wason's task suggests that inductive processes are likely involved in participants' responses. This effect cannot be attributed to memory recall.

In sum, the absence of effects on the Batman task suggests that it may be impractical (albeit not impossible since category abstraction and number of (card) categories did

influence response patterns on the abstract version of Wason's task; see *Results* section above) to speak of inductive processes in isolation from familiarity or background knowledge, which may function to alleviate the information load on working memory. Background knowledge may be a fundamental input variable for the operation of inductive processes. In the presence of both a conceptual load and an absence of background knowledge, inductive processes may fall back on a path of least resistance reasoning (see inductive-coherence model of reasoning presented in Chapter 2).

Study 2

In Study 1, conceptual distance and rule abstraction and number of (card) categories did not affect participants' responses on the Batman task. In contrast, rule abstraction and number of (card) categories did affect participants' responses to Wason's task. Given this pattern of results, I explored the reasons for such effects. As mentioned previously, the Batman task may have been too difficult, given its unusual storyline along with the quantity of information participants had to follow, to allow for tests of the variables. I reasoned that if participants were overwhelmed with the kind and amount of information presented in the task, they could have forgotten the purpose of the task and easily fallen into a single mode of responding as described in the inductive-coherence model. In such a mode of reasoning, participants likely failed to attend to specific variables in the task. In contrast, participants may have viewed Wason's task, which was presented following completion of the Batman task, as an altogether different and easier task. Given that Wason's task, in comparison requires much less information to keep in mind, participants may have modified their mode of responding. Given a different mode, rule abstraction and number of (card) categories might have been able to produce effects.

I tested this information-load hypothesis by incorporating two new manipulations into the existing Batman task. I expected these manipulations would make the Batman task easier for participants to solve or at least follow. First, I decided to *emphasize the instructions* in the Batman task so that participants would be reminded after every rule about the purpose of the task. In the original version (see Appendix 1), participants were

instructed only at the very beginning of the task that after every rule, they must choose the door(s) that would help them figure out if the rule is true or false. In the new version, participants were reminded after each of the 24 rules. I reasoned that even if the Batman task was difficult to solve, participants would always know what they had to do with each rule.

I also planned to make the Batman task easier by building in a *determinate* condition. In the original version of the task, participants were asked to test rules whose epistemic status they could not be sure of, that is, the rules could be either true or false. Some investigators believe that testing such indeterminate rules can be a difficult task and believe it is one reason why thematic versions, which require participants to assume the truth of the rule and seek violations to the rule, facilitate performance (Evans & Pollard, 1992; Wason, 1983). In the new Batman version, participants are asked to test rules whose epistemic status they can be sure of, namely, they are told the rules are *absolutely false* and that their task is *to select those door(s) that will show the rule to be false* (see Figure 4.5). The reason I decided to have absolutely false rules instead of absolutely true rules was so that the perspective of the story would not change. For example, given that in the story the Joker is responsible for creating the rules, it made little sense to have the Joker issue truthful rules. I could have had Batman issue absolutely true rules but then the perspective of the story would have had to change because now Batman was creating the rules so someone else would have had to test the rules (see original Batman task shown in Figure 4.2). Perspective change has been shown to alter performance on thematic versions of Wason's task (Gigerenzer & Hug, 1994). After participants finished one of the two versions of the Batman task, they were asked to indicate how confident they were about their performance on the Batman task. I asked participants about their confidence level so as to determine whether confidence level was predictive of performance in the Batman task. Finally, after indicating their confidence level, participants completed the same version of Wason's selection task used in Study 1.

Method

Participants. One hundred and twenty students from introductory psychology courses at the University of Alberta served as participants. All participants received class credit for collaborating, and none had training in logic.

Revised Batman Task. In Study 2, the (indeterminate) Batman task used in Study 1 was matched with the creation of a determinate Batman task (see Appendix 2) for all three rule abstraction and number of (card) category conditions. Hence, there were 6 rule abstraction and number of (card) category conditions all together; three conditions to again test the old indeterminate Batman task, and three conditions to now test the new determinate Batman task. The content of the rules used in the new determinate Batman task were identical to those of the old version. The only difference between the determinate and indeterminate conditions was in the epistemic status of the rules. In the determinate condition, participants were told that they knew the rules to be false and then instructed to *show the falsity* of the rules, whereas in the indeterminate condition, they were told that they did not know whether the rules were true or false and then instructed to *test the truthfulness or falsity* of the rules. In both the determinate and indeterminate tasks, instructions were clarified so that after presentation of each of the 24 rules participants were reminded of their task. After every rule in the new determinate version of the Batman task, participants were asked "Which door(s) would you choose in order to show that this rule is false?" (see Appendix 2). Likewise, after every rule in the old indeterminate Batman task, participants were asked "Which door(s) would you choose in order to figure out if this rule is true or false?" (see Appendix 2 but substitute the indeterminate instructions in place of the determinate instructions). After both Batman versions, participants were asked to rate their level of confidence in their task performance on a scale from 1 to 7, where 1 meant "not at all confident" and 7 meant "very confident." After indicating their confidence level, participants completed 6 trials of Wason's task as in Study 1. Instructions for Wason's task were kept standard as in Study 1. Participants followed the same procedure as in Study 1. Students completed the task in groups of 15 and finished the task in approximately 35 minutes.

Results

Overall frequency of responses. As in Study 1, participants across conditions continued to select the door representing the "p" card (16% to 31%), the doors representing the "p" and "q" cards (37% to 48%), and all the doors (7% to 8%) ($\chi2>42$, df>5, $p=0.00$). Only 4 to 11 percent of participants selected the correct doors, representing "p" and "not-q." Although participants' responses to Wason's task mirrored responses to the Batman task, more participants selected the correct cards in response to Wason' task than in response to the Batman task (12% to 16%).

Indeterminacy/Determinacy. Under the *indeterminate* Batman task context, neither conceptual distance nor rule abstraction and number of (card) categories had a significant effect on the *kind* or *number* of doors chosen to test the conditional rules. Recall that this was the same result obtained in Study 1. The lack of effects in the indeterminate Batman task goes against the hypothesis that clarification of the instructions should have simplified the task for participants.

Following the indeterminate Batman task, rule abstraction and number of (card) categories did not have an effect on participants' responses to Wason's task. Recall that in Study 1, this manipulation did have an effect on participants' responses to Wason's task; specifically, participants selected a greater number of doors to test the conditional rule as the conditions increased in information load.

Under the new *determinate* Batman task context, there were neither significant effects of conceptual distance nor significant effects of rule abstraction and number of (card) categories. The lack of effects in the determinate Batman task go against the hypothesis that both clarification of the instructions and the determinacy of the conditional rule should have simplified the task for participants. However, rule abstraction and number of (card) categories did have an effect on the *number* of doors participants selected to test the conditional rule in Wason's task following the determinate Batman task ($\chi2>10.52$; df=4; $p<.0325$ for 3 out of six trials; $\chi2>9.12$; df=4; $p<.0581$ for 2 out of six trials). As shown in Table 4.2, participants in condition 1 selected two or more doors to test the conditional rule at least 80 percent of the time. Similarly, participants in condition 2

selected two or more doors to test the conditional rule at least 90 percent of the time. In contrast, participants in condition 3 selected 2 or more doors only 55 percent of the time. The pattern of this result is similar to the one obtained in Study 1's (indeterminate) Batman task. Participants in conditions with a greater information load (conditions 1 and 2) selected a greater number of doors than participants in a condition with less of an information load. These results will be discussed in detail in the Discussion section.

Confidence levels. Under the *indeterminate* Batman task context, participants across the three conditions indicated similar levels of confidence in their performance. I was originally interested in whether a participant's confidence level influenced his or her performance on the Batman task. My expectation was that participants in conditions 1 and 2 should have lower confidence levels because of the increased number of (card) categories to consider and, hence, select a greater numbers of doors (Garner, 1962). They would select a greater number of doors because reduced confidence might prompt the investigation of more sources of evidence. However, given that the indeterminate Batman task did not appear to be simplified with the clarification of instructions, it is not surprising that participants did not differ across conditions in their confidence level. That is, the variable of number of (card) categories did not have an effect and, consequently, many participants responded alike.

In contrast, under the new *determinate* Batman task context, participants' across the three conditions indicated slightly different confidence levels in their performance ($\chi 2$=20.6857; df=12; p<.0552). As shown in Table 4.3, less than half of participants in conditions 1 and 2 indicated confidence levels above or equal to 5, whereas 55 percent of participants in condition 3 indicated confidence levels above or equal to 5. Moreover, 30 percent of participants in conditions 1 and 2 indicated they were somewhere in the middle (i.e., confidence level equal to 4). In contrast, only 10 percent of participants in condition 3 indicated they were somewhere in the middle. Similar percentages of participants in all three conditions indicated confidence levels below or equal to 3.

Discussion and Conclusion

An interesting pattern of effects emerged from this second study. First, there were no effects due to conceptual distance or rule abstraction and number of (card) categories under the *indeterminate* Batman task context (replication of Study 1). Unlike the result found in Study 1, I did not obtain an effect due to rule abstraction and number of (card) categories on Wason's task following the indeterminate Batman task. These results suggest that the clarification of instructions did not make the Batman task easier to comprehend and, in fact, may have served to emphasize the similarities between the Batman task and Wason's task. Participants overwhelmed by the indeterminate Batman task may have transferred their response patterns over to Wason's task, disrupting the variables of rule abstraction and number of (card) categories to influence participants' responses to this latter task.

Second, as with the indeterminate Batman task context, there were no effects due to conceptual distance or rule abstraction and number of (card) categories under the *determinate* Batman task context. In contrast, there was an effect of rule abstraction and number of (card) categories on Wason's task, following the determinate Batman task. The clarification of instructions in this case would not have served to make both tasks similar because the instructions given in the determinate Batman task required participants to assume the falsity of the rule and to test the rule by seeking instances that showed the rule to be false. In contrast, the instructions to Wason's task remained standard in so far as participants were told that they did not know whether the conditional rule was true or false. In this case, then, participants may have assumed each task to be unique and adopted a different response mode for each task. In the determinate Batman task, faced with the overwhelming amount of information and lack of background knowledge, participants may have adopted a single mode of responding; one in which rule abstraction and number of (card) categories failed to influence responses. In contrast, faced with the appearance of a different task, Wason's task, participants may have adopted a different mode of responding; one in which rule abstraction and number of (card) categories could influence responses.

Finally, participants across the three conditions associated with rule abstraction and number of (card) categories in the indeterminate Batman task context did not differ in their confidence levels. In contrast, participants across the three conditions in the determinate Batman task context differed in their confidence levels (chi-square approached significance). Collectively, these results suggest that rule abstraction and number of (card) categories had a minor effect on participants' confidence level as they responded to the Batman task. This result is understandable if we assume participants were still overwhelmed by the task and fell into a single mode of responding; that is, confidence may not have had much of a role if a heuristically based style of reasoning was used.

Nonetheless, two conclusions may be tentatively drawn from these results. First, it is clear from the results of both the first and the present study that the Batman task is a more difficult task for participants to solve than is even Wason's task. In both Studies 1 and 2, participants made a greater number of logically correct selections in response to Wason's task than to the Batman task. In addition, in both Studies 1 and 2, conceptual distance and rule abstraction along with number of (card) categories did not affect participants' responses to the Batman task, but rule abstraction and number of (card) categories did affect responses to Wason's task. I suspect that the Batman task in both studies was too difficult, prompting participants to revert to a single mode of responding as described in the inductive-coherence model presented in Chapter 2. The reasons why the Batman task may have been a more difficult task for participants to solve might include its unfamiliar context and its inability to cue participants' background knowledge. Although almost all participants are likely familiar with Batman, the super-hero and his associated escapades, they are likely completely unfamiliar with the context of the story; that is, a context where conditional rules with arbitrary terms must be tested. This task does not represent an everyday problem. Most participants do not have experience testing rules (whether false or indeterminate) about arbitrary objects. What the Batman task lacks is relevance for participants. The lack of effects on the Batman task strongly suggest that proponents of the memory-cuing hypothesis may be correct in stressing the role of familiarity (and

relevance) in the manifestation of thematic effects.

Second, in spite of the absence of effects on the Batman tasks, there is still evidence that participants do approach Wason's task inductively because rule abstraction and number of (card) categories did have an effect on participants' responses to Wason's task in both Studies 1 and 2. It appears that participants may have employed different reasoning modes in response to Wason's task. For example, for the Batman task, participants may have used a kind of heuristically based reasoning (leading to a single mode of responding) when confronted with such a cognitively demanding task lacking in meaningful content. In contrast, for Wason's task, participants may have changed their mode of responding in the face of a simpler cognitive task enough to allow inductive variables to influence their responses (see Tourangeu & Sternberg (1981) for a discussion of how the effects of an independent variable may be compromised by the presence of another variable).

Study 3

The motivation behind Studies 1 and 2 was to test the inductive hypothesis of deductive performance; that is, to obtain evidence that inductive processes underlie deductive performance as observed on Wason's task and associated thematic versions (i.e., the Batman task). In order to test this hypothesis without the variable of familiarity as a confounding factor (this variable is present in all thematic versions of Wason's task), I manipulated inductive variables that would not be expected to cue memory; for example, I manipulated conceptual distance, rule abstraction and number of (card) categories, and determinacy of the conditional rule. From the results obtained in Studies 1 and 2, rule abstraction and number of (card) categories was the only variable that influenced participants' performance and only on Wason's task. This result, which is not easily attributed to memory recall, is the main source of evidence thus far supporting the inductive hypothesis. The others variables I manipulated (i.e., conceptual distance and determinacy of the conditional rule) did not influence participants' performance on the Batman for reasons most likely attributed to task difficulty; this task demands that

participants keep a lot of information in mind without the advantage of cuing a conceptual frame in order to solve the task.

In order to test further the inductive hypothesis and to explore the factors that were interfering with participants' performance on the Batman task, I conducted think aloud interviews with participants as they attempted to solve the Batman task. To my knowledge, think aloud interviews have not heretofore been used to examine performance on thematic versions of Wason's task. Asking participants about their reasoning strategies as they solve the task seemed to be an obvious step in ascertaining the inductive (or deductive) nature of participants' reasoning.

Method

Participants. Twenty-two students from introductory psychology courses at the University of Alberta served as participants. All participants received class credit for collaborating, and none had training in logic.

Materials. The Batman tasks used in Studies 1 and 2 involved a paper-and-pencil format because I tested students in groups of 15 to 20 and so such a format was necessary for ease of administration. However, in Study 3, I was going to conduct think aloud interviews individually with each participant so a different Batman task format seemed warranted. For the purpose of Study 3, a miniature set of four rooms along with doors was built for participants to operate as they performed the Batman task. The miniature gadget, illustrated in Figure 4.6, was approximately 70 cm long, 7 cm wide at the top and 15 cm wide at the bottom. The doors were 8 cm long and 6 cm wide, with a little hook at the right of each door acting as a small door knob. Paper name plates were fastened on the outside of the doors with velcro, and paper room contents were placed inside each room without participants' knowledge.

Procedure. Students were interviewed individually in a small room. Eleven students read the *indeterminate* Batman task and another 11 students read the *determinate* task. Within each group of 11 students, 3-4 students were presented with condition 1 rules (i.e., rule abstraction and number of (card) category rules in the first condition), 3-4 students

with condition 2 rules, and 3-4 students with condition 3 rules. They were asked to read silently to themselves either the indeterminate or determinate version of the Batman task and then the investigator explained the task briefly and introduced students to the miniature set of doors they would be using during the task:

> So as you read from the story, your task is to test the truth or falsity (or to show the falsity) of the Joker's rules. You have not been given any rules yet; I will show you the first of his rules in a moment, but first let me explain the task in more detail. The Joker has made up some rules about how objects in the museum are arranged. You don't know whether these rules are true or false, however, you must find out if the rules are true or false (You know these rules are false and you must convince others of their falsity). What are you going to use as evidence to help you figure out if the rules are true or false (help you show that these rules are false). Well these museum rooms in front of you are going to help you out. These rooms hold potentially important information about the rule's truth or falsity (or falsity). This is the way you can use them: Given a rule, you should ask yourself this, 'Which door or doors would I have to open and/or close in order to find out if the rule is true or false (in order to show that the rule is false)?

The investigator went on to say:

> Why don't we see one of the Joker's rules. [Hand a rule to participant]. Please read the rule out loud. Given this rule, which door or doors would you have to open and/or close in order to find out if the rule is true or false (in order to show that the rule is false)? Point to the first door you would attend to and tell me what you are thinking as you select it.

Participants were asked about their thoughts for each door they selected and, afterward, for each door they did not select. Participants went through the same procedure for two rules. Only two rules from the 24 rules used in the group administration of the Batman task were chosen for participants to solve during the interview so as to (a) keep the interview under one hour, and (b) allow the interviewer to ask as many questions as needed and to allow participants sufficient time to answer (see Appendix 3 for the rules

used in each variable condition). Finally, participants were asked the following 8 questions designed to assess their overall impressions about the task and/or their strategies for solving the task:

(a) Did you find the story hard or easy to follow? Why?

(b) Did you find the task hard or easy to do? Why?

(c) How confident were you when you selected the doors in response to the first rule? Why?

(d) How confident were you when you selected the doors in response to the second rule? Why?

(e) When given a rule, were you initially prone to (1) believe, (2) disbelieve, or (3) neither believe it nor disbelieve it? Why?

(f) In general, how many door(s) do you think absolutely need to be checked in order to test the truth or falsity of the rule (in order to show the falsity of the rule)? Why?

(g) Is showing the rule to be true important? Why?

(h) Is showing the rule to be false important? Why?

Results

Overall frequency of responses. Although participants' performance was not "logically" correct, there were a number of interesting observations made from the interviews. First, participants responded alike to the rules irrespective of condition, that is, rule abstraction and number of (card) categories did not have an effect on performance. For this reason, interview results will not be presented by condition. Approximately 65 percent of participants selected the doors corresponding to "p" and "q" in order to test the first conditional rule. In contrast, only 36 percent of participants selected these doors in response to the second rule. The reason for the decline is most likely attributed to the feedback students obtained on their responses to the first rule. In particular, students were permitted to look at all the name plates and room contents after they had made their selections. Only two students (9%) selected the "logically" correct doors in response to

the first rule. This number, however, increased to 8 (36%) in response to the second rule. Finally, only one student selected the door corresponding to “p” in response to the first rule. This student then went on to select the “p” and “q” doors in response to the second rule. Interestingly, there were 4 (18%) students who selected the door corresponding to “not-p” either by itself or in combination with another selection in response to the first rule. This number increased to six (27%) in response to the second rule; only two students selected this door consistently in response to the first and second rule. Only 1 student selected all doors in response to a rule.

A wealth of data was obtained from the interviews. Some of the most interesting information, however, came from the set of eight questions asked of students at the end of the task. It is primarily here that one gets a sense of how students approached the Batman task.

Did you find the story hard or easy to follow? Why? In response to this question, 50 percent of participants said the story was “difficult” or they “had to read it twice to get it.” The other 50 percent said the story was easy, but when asked why, they typically cited good grammar or neat format as reasons. One student said that “it was easy, but then again I’m typically good with games and puzzles.”

Did you find the task hard or easy to do? Why? In response to the second question, 64 percent of students claimed the task was hard or tricky. The other 35 percent commented they thought it was easy at first, but then realized it was more difficult than first imagined. A comment made by many of these students was “the task is hard because there is so much thought that needs to go into it. It’s difficult because you have to think about all the doors.” In short, all students found the task hard. Some realized this right away, while others did after they received feedback about their responses to the first rule.

How confident were you when you selected the doors in response to the first and second rule? Why? In response to this question, 55 percent of students said their confidence grew as they went from responding to rule 1 to rule 2. A typical reason for this increase in confidence involved having a better idea about the task especially after participants had the opportunity to look at all doors after responding to the first rule.

Twenty-seven percent of students indicated that their confidence had suffered as they went from rule 1 to rule 2. A typical reason for the decrease in confidence involved the awareness that all doors could potentially be important to testing (or showing the falsity) of the rule. Finally, 18 percent of students indicated no change in their confidence as they went from rule 1 to rule 2. One student mentioned that her confidence did not grow or lessen because the feedback obtained from opening and/or closing all the doors after the first rule was not helpful. Another student mentioned that the level of difficulty had remained the same for both rule 1 and 2.

When given a rule, were you initially prone to (1) believe, (2) disbelieve, or (3) neither believe it nor disbelieve it? Why? This question revealed interesting answers. Approximately 60 percent of students said that they were initially prone to believe the rules in spite of knowing from the story that either the rules might be true or false or absolutely false. The two reasons given for the propensity to believe the rules included (a) if...then statements are usually true so it is hard to bypass the “urge” to believe it, and (b) assuming the truth of the rule functioned as a starting point from which doors are selected to test the rule (many of these students indicated that testing the truth or falsity of the rule was important but it was easier to assume the rule to be true, check for positive instances, and then watch for deviations to the rule). Twenty-seven percent of students said they neither believed nor disbelieved the rule. Finally, one student claimed that her initial reaction was to disbelieve the rules but did assume the truth of the rules as a starting point from which to select doors.

In general, how many door(s) do you think absolutely need to be checked in order to test the truth or falsity of the rule (in order to show the falsity of the rule)? Why? Approximately 40 percent of students said that it was difficult to say precisely how many doors were absolutely necessary to test the truth or falsity (or show the falsity) of the rules because whether one chose the right door to test (or to falsify) the rule was partly a matter of chance. Most agreed that one should keep opening doors until the falsifying instance was found--even if this meant selecting all doors. Most students, however, did not open all 4 doors during the task. When asked about the discrepancy, some students said that the

feedback they had obtained about their responses had taught them that opening all doors would be a useful strategy to use in future questions. Other students mentioned that in order to be practical, they had first chosen the doors that matched the rule. If these selections had only confirmed the rule, they would have opened the other doors "just to make sure it was really false." Forty-six percent of students said two doors were sufficient to prove the rule's truth or falsity (or show its falsity), while 9 percent of students claimed that one would have to open and/or close 3 doors to be sure of the rule's truth or falsity (or falsity). Finally, one student said that one door was sufficient.

Is showing the rule to be false or true important? Why? Approximately 30 percent of students admitted that, in spite of the story, proving the rule to be true motivated their actions more than proving the rule to be false. Fifty-two percent of students claimed that they wanted to either prove the rule true or false but that they assumed the rule to be true as a starting point to guide their selections. Finally, 18 percent of students claimed that proving the rule to be false was most important and, interestingly, many of these students cited University or high school courses as reasons for thinking this way. For example, one student said "I took a course where my prof said that most people try and prove theories, but you should try to disprove them."

Discussion and Conclusion

Think aloud interviews were conducted with students as they attempted to solve the Batman task in order to find out the factors that were interfering with participants' performance on the task and also to determine whether inductive or deductive strategies played a role in their performance. Preliminary answers to both these questions were obtained and should guide future studies designed to test the inductive hypothesis. First, it is interesting that few of participants' responses suggested the spontaneous use of deductive or logical rules. Of the two students who provided correct selections to the first conditional rule, only one student appeared to be employing spontaneous logical strategies. The second student said he knew how to do the task because one of his professors had explained briefly the standard method of hypothesis testing (i.e., seek

falsifying evidence). Another 5 students were logically correct in their responses only after they had received feedback to their responses to the first conditional rule. Of these 5 students, two mentioned that they had either read or recalled proper methods of hypothesis-testing. In general, although participants did exhibit logical methods of hypothesis-testing, it is unclear whether these procedures reflect spontaneous deductive strategies or learned performance because (a) only one out of 22 participants used such a procedure spontaneously in response to the first rule, and (b) the 5 others who also used such a procedure only did so after they received feedback about their "inductive" first round performance. If logical rules are applied in problem-solving domains, it appears that such rules are learned formally in school.

Second, the responses obtained from the interviews suggest that the Batman task is difficult for participants to solve both because of its context and the nature of the task. It is reasonable to assume that participants in Studies 1 and 2 likely had similar impressions to those of participants in Study 3. Students in the first two studies may have chosen to not expend too much effort on such a difficult task. The lesson learned from the Batman task is that any thematic context will not do; relevant background knowledge appears to be necessary in order to build a conceptual frame around the context of the task. Finally, the interviews also suggest that inductive strategies are motivating many of students' responses. Many students learned from the first rule to the second, changing their responses accordingly. Moreover, many students attempted to solve the task by assuming the truth of the rule because, according to them, based on their everyday experiences with conditional rules, it is easier to test rules by assuming their truth and then to seek deviations. These results are to be expected from a reasoning system that exploits background knowledge in problem solving pursuits, and can change strategies depending on additional information.

Finally, from the results presented in this chapter, it is possible to draw some general conclusions regarding the role of both inductive and deductive reasoning. First, participants do appear to approach Wason's task inductively given that rule abstraction and number of (card) categories did affect participants' responses to the task. Second, the

goal of separating familiarity from induction in future tests of the inductive hypothesis may be impractical if inductive processes are intimately dependent on background knowledge for their operation (Holland et al., 1986). I would argue that believing that inductive processes can be isolated from the influence of familiarity is a (vestige) assumption that can be traced back to the formal or deductive hypothesis; that is, the content of the task is somehow secondary to the structure or syntax of the task. There is overwhelming evidence that content plays a fundamental role in inductive performance (e.g., Cummins, 1995; Cummins et al., 1991; Cosmides, 1989; Griggs & Cox, 1982; Holland et al., 1986). Third, the results presented in this chapter suggest that participants are knowledgeable about deductive reasoning (or performance) but that this is not their first choice when reasoning about a task. The students I tested in the third study clearly indicated an inductive disposition to solving the Batman task. Some students did apply deductive strategies but only after they received feedback and attempted to recall what their professor had taught in class. Deduction is a vital method of reasoning that appears to require intensive instruction if it is to be done well and fluidly. Anecdotally, it is of little wonder that those domains that depend most on deductive, logical reasoning (e.g., computing, engineering, mathematics, physics etc.) are also those in which students need the most formal training.

Chapter 5
General Conclusions

At the beginning of this paper I set out to present an explicitly informal account of participants' performance on traditional formal reasoning tasks. I argued that this presentation required examining participants' formal performance within an everyday problem solving framework. Within this framework, participants' underlying reasoning processes were understood to be inductive and able to produce either deductive or inductive strategies. Inductive strategies, moreover, were expected to characterize the kind of unrehearsed reasoning measured by standard formal tasks used in psychological experiments. Consequently, I argued, biases and errors occur when people attempt to solve formal tasks with "inductive" strategies.

In Chapter 2, I proposed the inductive-coherence model of reasoning and laid out the ground work (or rationale) for thinking of formal task performance as a largely inductive enterprise. The reasons for supporting the inductive-coherence model included the nature of the problems people must solve in everyday life, and the incompatibility between the data obtained from formal tasks and formal explanations. In Chapter 3, I presented the kind of inductive architecture that might underlie the inductive-coherence model. To this end, I demonstrated how a PDP network could learn to solve a traditional formal task such as Wason's task by means of pattern classification (a form of induction). Finally, in Chapter 4, I wanted to further test the inductive hypothesis and also determine whether this hypothesis could be meaningfully separated from the memory-cuing or "familiarity" hypothesis. Although in the inductive-coherence model a real separation between the factors of inductive reasoning and background knowledge (i.e., familiarity) is impractical because both factors are intimately intertwined, belief of this separation is suggested by a number of theorists (e.g., Cosmides, 1989). In Chapter 4, moreover, I demonstrated that manipulating specific inductive variables (e.g., conceptual distance, rule abstraction and number of (card) categories, and the determinacy of the rule in the thematic Batman task) did not have an effect on participants' performance. Although rule abstraction and

number of (card) categories had an effect on participants' responses to Wason's task, the other two inductive variables did not significantly influence performance on the Batman task. The results obtained from these studies suggest that familiarity may not be practically separated from inductive reasoning processes.

Finally, In order to investigate the strategies participants use to solve formal tasks, participants in Study 3 were interviewed individually and asked to think aloud as they performed the Batman task. Results from this last study supported the hypothesis that (a) people predominantly employ inductive strategies to solve a thematic version of Wason's task (i.e., the Batman task), (b) deductive strategies are not spontaneously used to solve the task, and (c) the Batman task was indeed difficult for participants to solve, but not because it was grammatically incorrect or incoherent, but because it lacked a fundamental variable. The Batman task lacked the elements found in most everyday problems, which render them interpretable and interesting. The Batman task lacked familiarity and relevance. Although the task "made sense" to participants as a story or game, the task was not any everyday problem typically encountered. As a result, participants were unable to impose background knowledge to solve the task, and make sense of it at a deep level.

The objective now is to try and test induction within the boundaries of familiar contexts. To this end, a careful analysis of the variables that motivate in-depth reasoning in everyday contexts (e.g., cost/benefit analysis) need to be investigated.

Initially the claim was made that we could gain a greater understanding of reasoning if we looked at it from a problem solving perspective. By looking at reasoning from this perspective, it is possible to appreciate that errors observed on formal tasks are really not errors at all. Participants' common responses on formal tasks are reasonable responses given the features or constraints of inductive processes. Specifically, inductive reasoning is highly dependent on background beliefs in order for the problem space to be minimized. This minimization allows inductive processes to yield useful and quick solutions to everyday problems. Often times, alternate solutions will not be explored to problems because the most available solution is adequate (Rescher, 1980).

Because inductive reasoning is so dependent on background knowledge, our

tendencies will also reflect a healthy conservatism in revising our beliefs and questioning known beliefs. Conservatism is adaptive in so far as it preserves useful beliefs in spite of (weak) contradictory evidence against them. This conservatism can also be maladaptive, however, if individuals need to make high stake decisions, in which exploring all possible alternative solutions is critical. Such decisions seem to pervade the highly complex everyday life that people currently live in today's information age. For this reason, as researchers we must concentrate on identifying the conditions and variables that will facilitate sound *inductive* reasoning. In order to accomplish this, we cannot continue to entertain beliefs about the (formal) underlying processes of reasoning, which empirical evidence fails to support. It is necessary now to shift our attention to the study of informal reasoning processes with real, everyday tasks.

In sum, one lesson that emerges from this paper is that positing an explicitly informal account of formal task performance such as is done with the inductive-coherence model can offer a new perspective on performance previously characterized as biased. A second lesson is that by separating the model from the architecture, it was possible to show that an inductive architecture can generate formal performance. This is further evidence for an explicitly informal account of formal task performance. Finally, the explicitly informal account is supported with evidence that "familiarity" or background knowledge may not be meaningfully separated from inductive processes in participants' formal task performance. This is exactly what one would expect from an inductive system that fundamentally relies on a person's background knowledge to generate solutions to tasks. Ultimately, knowledge about the boundary conditions associated with human reasoning is key to understanding and facilitating not only inductive performance but also deductive performance.

Table 3.1

Binary Coding of Conditional Rule Types and Cards Used to Train Networks and Generate Input Patterns

Rules: Text Form	Rules: Binary Form
If vowel then even number	0 0 1 0
If vowel then odd number	0 0 1 1
If consonant then even number	0 1 1 0
If consonant then odd number	0 1 1 1
If even number then vowel	1 0 0 0
If even number then consonant	1 0 0 1
If odd number then vowel	1 1 0 0
If odd number then consonant	1 1 0 1
Cards: Text Form	**Cards: Binary Form**
A	0 0 0
E	0 0 1
J	0 1 0
K	0 1 1
4	1 0 0
6	1 0 1
5	1 1 0
7	1 1 1

Table 3.2

Definite Features for Bands From Hidden Units of Network 1

Hidden Unit	Band Label	Definite Features	N[1]
0	A	I0 ≠ I2, I0 ≠ I10, I2 = I10	1152
0	B	I0 ≠ I2, I1 ≠ I11	1152
0	C	I0 ≠ I2, I0 = I10, I1 = I11, I2 ≠ I10	768
1	A	I0 ≠ I2, I0 ≠ I7, I2 = I7	1152
1	B	I0 ≠ I2, I1 ≠ I8	1152
1	C	I0 ≠ I2, I0 = I7, I1 = I8 I2 ≠ I7	768
2	A	I0 ≠ I2, I0 ≠ I4, I2 = I4	1152
2	B	I0 ≠ I2, I1 ≠ I5	1152
2	C	I0 ≠ I2, I0 = I4, I1 = I5 I2 ≠ I4	768

Note. "≠" indicates a perfectly negative correlation between input units; "=" indicates a perfectly positive correlation between input units.

[1]N=Number of patterns falling in each band.

Table 3.3

Correlation Among Hidden Units When Cards 1 and 2 are Selected

Card 1								
	H0	H1	H2	H3	H4	H5	H6	H7
H0	1							
H1	0.09	1						
H2	0.16	-0.32	1					
H3	0	-0.11	0	1				
H4	0.02	-0.31	0.02	-0.19	1			
H5	0	-0.24	-0.28	0.154	0.18	1		
H6	0.542	-0.1	-0.1	-0.13	0.04	0.16	1	
H7	0	0.09	0	-0.99	0.212	0	0.144	1

Card 2								
	H0	H1	H2	H3	H4	H5	H6	H7
H0	1							
H1	0	1						
H2	0.04	-0.32	1					
H3	-0.16	-0.25	-0.21	1				
H4	0.06	-0.31	0.02	0.03	1			
H5	-0.18	-0.24	-0.28	0.02	0.181	1		
H6	0.996	0	0.02	-0.16	0.07	-0.1	1	
H7	0.191	-0.11	-0.13	-0.34	0.197	0.105	0.21	1

Table 3.4

Correlation Among Hidden Units When Cards 3 and 4 are Selected

Card 3								
	H0	H1	H2	H3	H4	H5	H6	H7
H0	1							
H1	0.09	1						
H2	0	-0.24	1					
H3	0.307	-0.25	0.03	1				
H4	0.02	-0.31	0.18	0.03	1			
H5	0	-0.24	1	0.03	0.18	1		
H6	0.541	-0.1	0.161	0.03	0.04	0.16	1	
H7	-0.1	-0.11	0.106	-0.34	0.196	0.105	-0.22	1

Card 4								
	H0	H1	H2	H3	H4	H5	H6	H7
H0	1							
H1	-0.1	1						
H2	0.16	-0.21	1					
H3	0.309	-0.1	-0.21	1				
H4	-0.1	0.814	-0.1	-0.15	1			
H5	0	0.748	-0.28	0.03	0.224	1		
H6	0.541	0.07	-0.1	0.03	0	0.159	1	
H7	-0.1	0.173	-0.13	-0.34	0.163	0.106	-0.22	1

Table 3.5

Definite Features for Bands From Hidden Units 0 and 6 of Network 2

H0-Bands	Definite Features	N[1]	H6-Bands	Definite Features	N
A	I0 ≠ I1, I0 ≠ I2, I0 = I3 I0 ≠ I7, I0 ≠ I8, I1 = I2, I1 ≠ I3, I1 = I7, I1 = I8 I2 ≠ I3, I2 = I7, I2 = I8 I3 ≠ I7, I3 ≠ I8, I7 = I8	192	A	I0 ≠ I2	1344
			B	I0 ≠ I2, I1 ≠ I3	576
B	I0 = I1, I0 ≠ I2, I0 = I3 I0 ≠ I7, I0 ≠ I8, I1 ≠ I2, I1 = I3, I1 ≠ I7, I1 ≠ I8 I2 ≠ I3, I2 = I7, I2 = I8 I3 ≠ I7, I3 ≠ I8, I7 = I8	192	C	I0 = I1, I0 ≠ I2, I0 ≠ I3 I0 = I7, I0 ≠ I8, I1 ≠ I2 I1 ≠ I3, I1 = I7, I1 ≠ I8 I2 = I3, I2 ≠ I7, I2 = I8 I3 ≠ I7, I3 = I8, I7.≠ I8	192
C	I0 ≠ I2, I0 = I7, I1 ≠ I3 I1 = I8, I2 ≠ I7, I3 ≠ I8	384	D	I0 = I1, I0 ≠ I2, I0 ≠ I3 I0 ≠ I7, I0 = I8, I1 ≠ I2 I1 ≠ I3, I1 ≠ I7, I1 = I8 I2 = I3, I2 = I7, I2 ≠ I8 I3 = I7, I3 ≠ I8, I7 ≠ I8	192
D	I0 ≠ I2, I1 = I3	576			
E	I0 = I1, I0 ≠ I2, I0 ≠ I3 I0 ≠ I7, I0 = I8, I1 ≠ I2, I1 ≠ I3, I1 ≠ I7, I1 = I8 I2 = I3, I2 = I7, I2 ≠ I8 I3 = I7, I3 ≠ I8, I7 ≠ I8	192	E	I0 ≠ I1, I0 ≠ I2, I0 = I3 I0 = I7, I0 = I8, I1 = I2 I1 ≠ I3, I1 ≠ I7, I1 ≠ I8 I2 ≠ I3, I2 ≠ I7, I2 ≠ I8 I3 = I7, I3 = I8, I7 = I8	192
F	I0 ≠ I1, I0 ≠ I2, I0 = I3 I0 ≠ I7, I0 = I8, I1 = I2 I1 ≠ I3, I1 = I7, I1 ≠ I8 I2 ≠ I3, I2 = I7, I2 ≠ I8 I3 ≠ I7, I3 = I8, I7 ≠ I8	192			
G	I0 = I1, I0 ≠ I2, I0 ≠ I3 I0 ≠ I7, I0 ≠ I8, I1 ≠ I2 I1 ≠ I3, I1 ≠ I7, I1 ≠ I8 I2 = I3, I2 = I7, I2 = I8 I3 = I7, I3 = I8, I7 = I8	192			
H	I0 ≠ I2, I1 = I3	768			
I	I0 ≠ I2, I0 = I7, I1 ≠ I3 I1 ≠ I8, I2 ≠ I7, I3 = I8	384			

Note. "≠" indicates a perfectly negative correlation between input units; "=" indicates a perfectly positive correlation between input units.

[1]N=Number of patterns falling in each band.

Table 3.6

Definite Features for Bands From Hidden Units 1 and 4 of Network 2

H1-Bands	Definite Features	N[1]	H4-Bands	Definite Features	N
A	I0 ≠ I2, I1 ≠ I14	960	A	I0 = I1, I0 ≠ I2, I0 ≠ I13, I0 = I14, I1 ≠ I2, I1 ≠ I13, I1 = I14, I2 = I13, I2 ≠ I14, I13 ≠ I14	384
B	I0 ≠ I1, I0 ≠ I2, I0 ≠ I3, I1 = I2 , I1 = I3, I2 = I3 I13 = I14	384			
C	I0 = I1, I0 ≠ I2, I0 = I3, I0 = I13, I0 ≠ I14, I1 ≠ I2 I1 = I3, I1 ≠ I13 I1 ≠ I4, I2 ≠ I3, I2 ≠ I13, I2 =I14 I3 = I13, I3 ≠ I14, I13 ≠ I14	192	B	I0 = I1, I0 ≠ I2, I0 = I3, I0 ≠ I13, I0 ≠ I14, I1 ≠ I2, I1 = I3, I1 ≠ I13, I1 ≠ I14, I2 ≠ I3, I2 = I13, I2 = I14, I3 ≠ I13, I3 ≠ I14, I13 = I14	192
D	I0 = I1, I0 ≠ I2, I0 = I3, I0 ≠ I13, I0 = I14, I1 ≠ I2, I1 ≠ I3, I1 ≠ I13 I1 = I14, I2 = I3, I2 = I13, I2 ≠ I14, I3 = I13, I3 ≠ I14, I13 ≠ I14	192	C	I0 = I1, I0 ≠ I2, I0 ≠ I3, I0 ≠ I13, I0 ≠ I14, I1 ≠ I2, I1 ≠ I3, I1 ≠ I13, I1 ≠ I14, I2 = I3, I2 = I13, I2 = I14, I3 = I13, I3 = I14, I13 = I14	192
E	I0 = I1, I0 ≠ I2, I0 = I3, I0 = I13, I0 = I14, I1 ≠ I2, I1 = I3, I1 = I13, I1 = I14, I2 ≠ I3, I2 ≠ I13, I2 ≠ I14, I3 = I13, I3 = I14, I13 = I14	192	D	I0 ≠ I1, I0 ≠ I2, I0 = I3, I0 ≠ I13, I0 = I14, I1 = I2, I1 ≠ I3, I1 = I13, I1 ≠ I14, I2 ≠ I3, I2 = I13, I2 ≠ I14, I3 ≠ I13, I3 = I14, I13 ≠ I14	192
F	I0 ≠ I1, I0 ≠ I2, I0 ≠ I3, I0 ≠ I13, I0 = I14, I1 = I2, I1 = I3, I1 = I13, I1 ≠ I14, I2 = I3, I2 = I13, I2 ≠ I14, I3 = I13, I3 ≠ I14, I13 ≠ I14	192	E	I0 ≠ I1, I0 ≠ I2, I0 ≠ I3, I0 ≠ I13, I0 = I14, I1 = I2, I1 = I3, I1 = I13, I1 ≠ I14, I2 = I3, I2 = I13, I2 ≠ I14, I3 = I13, I3 ≠ I14, I13 ≠ I14	192
G	I0 ≠ I1, I0 ≠ I2, I0 ≠ I3, I0 = I13, I0 ≠ I14, I1 = I2, I1 = I3, I1 ≠ I13 I1 = I14, I2 = I3, I2 ≠ I13, I2 = I14, I3 ≠ I13, I3 = I14, I13 ≠ I14	192	F	I0 ≠ I1, I0 ≠ I2, I0 ≠ I3, I0 = I13, I0 ≠ I14, I1 = I2, I1 = I3, I1 ≠ I13, I1 = I14, I2 = I3, I2 ≠ I13, I2 = I14, I3 ≠ I13, I3 = I14, I13 ≠ I14	192
H	I0 ≠ I1, I0 ≠ I2, I0 = I3, I0 = I13, I0 ≠ I14, I1 = I2, I1 ≠ I3, I1 ≠ I13, I1 = I14, I2 ≠ I3, I2 ≠ I13, I2 = I14, I3 = I13, I3 ≠ I14, I13 ≠ I14	192	G	I0 ≠ I1, I0 ≠ I2, I0 = I3, I0 = I13, I0 ≠ I14, I1 = I2, I1 ≠ I3, I1 ≠ I13, I1 = I14, I2 ≠ I3, I2 ≠ I13, I2 = I14, I3 = I13, I3 ≠ I14, I13 ≠ I14	192
I	I0 ≠ I1, I0 ≠ I2, I0 = I3, I0 ≠ I13, I0 ≠ I14, I1 = I2, I1 ≠ I3, I1 = I13, I1 = I14, I2 ≠ I3, I2 = I13, I2 = I14, I3 ≠ I13, I3 ≠ I14, I13 = I14	192	H	I0 = I1, I0 ≠ I2, I0 = I3, I0 = I13, I0 = I14, I1 ≠ I2, I1 = I3, I1 = I13, I1 = I14, I2 ≠ I3, I2 ≠ I13, I2 ≠ I14, I3 = I13, I3 = I14, I13 = I14	192
J	I0 = I1, I0 ≠ I2, I0 ≠ I3, I0 = I13, I0 = I14, I1 ≠ I2, I1 ≠ I3, I1 = I13, I1 = I14, I2 = I3, I2 ≠ I13, I2 ≠ I14, I3 ≠ I13, I3 ≠ I14, I13 = I14	192	I	I0 ≠ I1, I0 ≠ I2, I0 = I3, I0 ≠ I13, I0 ≠ I14, I1 = I2, I1 ≠ I3, I1 = I13, I1 = I14, I2 ≠ I3, I2 = I13, I2 = I14, I3 ≠ I13, I3 ≠ I14, I13 = I14	192
K	I0 = I1, I0 ≠ I2, I0 = I3, I0 ≠ I13, I0 = I14, I1 ≠ I2, I1 = I3, I1 ≠ I13, I1 = I14, I2 ≠ I3, I2 = I13, I2 ≠ I14, I3 ≠ I13, I3 = I14, I13 ≠ I14	192	J	I0 = I1, I0 ≠ I2, I0 ≠ I3, I0 = I13, I0 = I14, I1 ≠ I2, I1 ≠ I3, I1 = I13, I1 = I14, I2 = I3, I2 ≠ I13, I2 ≠ I14, I3 ≠ I13, I3 ≠ I14, I13 = I14	192
			K	I0 ≠ I1, I0 ≠ I2, I0 ≠ I3, I0 ≠ I13, I0 ≠ I14, I1 = I2, I1 = I3, I1 = I13, I1 = I14, I2 = I3 I2 = I13, I2 = I14, I3 = I13, I3 = I14, I13 = I14	192
			L	I0 ≠ I1, I0 ≠ I2, I0 ≠ I3, I0 = I13, I0 = I14, I1 = I2, I1 = I3, I1 ≠ I13, I1 ≠ I14, I2 = I3 I2 ≠ I13, I2 ≠ I14, I3 ≠ I13, I3 ≠ I14, I13 = I14	192
			M	I0 ≠ I2, I0 = I3, I1 ≠ I14, I2 ≠ I13	576

Note. "≠" indicates a perfectly negative correlation between input units; "=" indicates a perfectly positive correlation between input units. [1]N=Number of patterns falling in each band.

Table 3.7

Definite Features for Bands From Hidden Units 2 and 5 of Network 2

H2-Bands	Definite Features	N[1]	H5-Bands	Definite Features	N
A	I0 ≠ I2, I1 ≠ I3, I1 ≠ I10, I1 ≠ I11 I3 = I10, I3 = I11, 10 = I11	384	A	I0 = I1, I0 ≠ I2, I0 = I3 I1 ≠ I2, I1 = I3, I2 ≠ I3	768
B	I0 ≠ I2, I1 ≠ I3, I1 = I10, I1 ≠ I11 I3 ≠ I10, I3 = I11, I10 ≠ I11	384	B	I0 ≠ I1, I0 ≠ I2, I0 ≠ I3 I1 = I2, I1 = I3, I2 = I3	768
C	I0 ≠ I2, I1 = I3, I10 = I11	768	C	I0 ≠ I2, I1 ≠ I3	1536
D	I0 ≠ I2, I1 = I3, I10 ≠ I11	768			
E	I0 ≠ I2, I1 ≠ I3, I1 = I11, I3 ≠ I11	768			

Note. "≠" indicates a perfectly negative correlation between input units; "=" indicates a perfectly positive correlation between input units.

[1]N=Number of patterns falling in each band.

Table 3.8

Definite Features for Bands From Hidden Units 3 and 7 of Network 2

H3-Bands	Definite Features	N[1]	H7-Bands	Definite Features	N
A	I0 ≠ I2	1536	A	I0 ≠ I2	1152
B	I0 ≠ I2	768	B	I0 = I1, I0 ≠ I2, I0 = I3, I0 = I4, I0 ≠ I5, I1 ≠ I2, I1 = I3, I1 = I4, I1 ≠ I5, I2 ≠ I3, I2 ≠ I4, I2 = I5 I3 = I4, I3 ≠ I5, I4 ≠ I5	192
C	I0 = I1, I0 ≠ I2, I0 = I3, I0 = I4, I0 = I5, I1 ≠ I2, I1 = I3, I1 = I4, I1 = I5, I2 ≠ I3, I2 ≠ I4, I2 ≠ I5 I3 = I4, I3 = I5, I4 = I5	192	C	I0 ≠ I1, I0 ≠ I2, I0 = I3, I0 ≠ I4, I0 ≠ I5, I1 = I2, I1 ≠ I3, I1 = I4, I1 = I5,, I2 ≠ I3, I2 = I4, I2 = I5 I3 ≠ I4, I3 ≠ I5, I4 = I5	192
D	I0 = I1, I0 ≠ I2, I0 ≠ I3, I0 ≠ I4, I0 = I5, I1 ≠ I2, I1 ≠ I3, I1 ≠ I4, I1 = I5, I2 = I3, I2 = I4, I2 ≠ I5 I3 = I4, I3 ≠ I5, I4 ≠ I5	192	D	I0 = I1, I0 ≠ I2, I0 ≠ I3, I0 ≠ I4 I0 = I5, I1 ≠ I2, I1 ≠ I3, I1 ≠ I4 I1 = I5, I2 = I3, I2 = I4, I2 ≠ I5 I3 = I4, I3 ≠ I5, I4 ≠ I5	192
E	I0 ≠ I1, I0 ≠ I2, I0 = I3, I0 ≠ I4, I0 ≠ I5, I1 = I2, I1 ≠ I3, I1 = I4, I1 = I5, I2 ≠ I3, I2 = I4, I2 = I5 I3 ≠ I4, I3 ≠ I5, I4 = I5	192	E	I0 = I1, I0 ≠ I2, I0 ≠ I3, I0 ≠ I4 I0 ≠ I5, I1 ≠ I2, I1 ≠ I3, I1 ≠ I4 I1 ≠ I5, I2 = I3, I2 = I4, I2 = I5 I3 = I4, I3 = I5, I4 = I5	192
F	I0 = I1, I0 ≠ I2, I0 ≠ I3, I0 = I4, I0 = I5, I1 ≠ I2, I1 ≠ I3, I1 = I4, I1 = I5, I2 = I3, I2 ≠ I4, I2 ≠ I5 I3 ≠ I4, I3 ≠ I5, I4 = I5	192	F	I0 = I1, I0 ≠ I2, I0 ≠ I3, I0 = I4 I0 ≠ I5, I1 ≠ I2, I1 ≠ I3, I1 = I4 I1 ≠ I5, I2 = I3, I2 ≠ I4, I2 = I5 I3 ≠ I4, I3 = I5, I4 ≠ I5	192
			G	I0 ≠ I1, I0 ≠ I2, I0 = I3, I0 ≠ I4 I0 = I5, I1 = I2, I1 ≠ I3, I1 = I4 I1 ≠ I5, I2 ≠ I3, I2 = I4, I2 ≠ I5 I3 ≠ I4, I3 = I5, I4 ≠ I5	192
			H	I0 ≠ I1, I0 ≠ I2, I0 = I3, I0 = I4 I0 ≠ I5, I1 = I2, I1 ≠ I3, I1 ≠ I4 I1 = I5, I2 ≠ I3, I2 ≠ I4, I2 = I5 I3 = I4, I3 ≠ I5, I4 ≠ I5	192
			I	I0 = I1, I0 ≠ I2, I0 = I3, I0 ≠ I4 I0 ≠ I5, I1 ≠ I2, I1 = I3, I1 ≠ I4 I1 ≠ I5, I2 ≠ I3, I2 = I4, I2 = I5 I3 ≠ I4, I3 ≠ I5, I4 = I5	192
			J	I0 ≠ I1, I0 ≠ I2, I0 ≠ I3, I1 = I2 I1 = I3, I2 = I3, I4 ≠ I5	384

Note. "≠" indicates a perfectly negative correlation between input units; "=" indicates a perfectly positive correlation between input units. [1]N=Number of patterns falling in each band.

Table 4.1

Frequency of Participants in Each Category Abstraction Condition Who Selected One, Two, or More Than Three Doors to Test Rules

	Condition 1		Condition 2		Condition 3	
	Batman	Wason	Batman	Wason	Batman	Wason
One door selected	4	2	5	4	8	8
Two doors selected	12	16	9	10	9	10
>Three doors selected	3	1	5	5	2	1
Total	19	19	19	19	19	19

Note. There were only differences in the contingency table associated with performance on Wason's task ($\chi 2=10.568$, df=4, $p=0.05$).

Table 4.2

Frequency of Participants by Batman Version and Category Abstraction Who Selected One, Two, or More Than Three Doors to Test Rules

	Indeterminate Batman Task					
	Condition 1		Condition 2		Condition 3	
	Batman	Wason	Batman	Wason	Batman	Wason
One door selected	6	5	7	3	6	6
Two doors selected	13	12	12	12	12	11
>Three doors selected	1	3	1	5	2	3
Total	20	20	20	20	20	20
	Determinate Batman Task					
	Condition 1		Condition 2		Condition 3	
	Batman	Wason	Batman	Wason	Batman	Wason
One door selected	5	3	3	1	7	9
Two doors selected	12	14	13	15	11	8
>Three doors selected	3	3	4	4	2	3
Total	20	20	20	20	20	20

Note. There were only differences in the contingency table associated with performance on Wason's task in the Determinate condition ($\chi 2=10.534$, df=4, $p=0.05$).

Table 4.3

Frequency of Participants at Each of 7 Confidence Levels by Abstraction and Batman Task Context

		Indeterminate Batman Task Context			
		Condition 1	Condition 2	Condition 3	Total
Confidence Level	1	2	3	2	7
	2	1	3	5	9
	3	3	3	6	12
	4	2	4	2	8
	5	6	4	1	11
	6	3	3	3	9
	7	3	0	1	4
	Total	20	20	20	60

$\chi 2=12.4069$, df=12, $p=.4136$

		Determinate Batman Task Context			
		Condition 1	Condition 2	Condition 3	Total
Confidence Level	1	1	1	3	5
	2	2	7	1	10
	3	2	0	3	5
	4	6	6	2	14
	5	6	1	3	10
	6	3	4	5	12
	7	0	1	3	4
	Total	20	20	20	60

$\chi 2=20.6857$, df=12, $p=.0552$

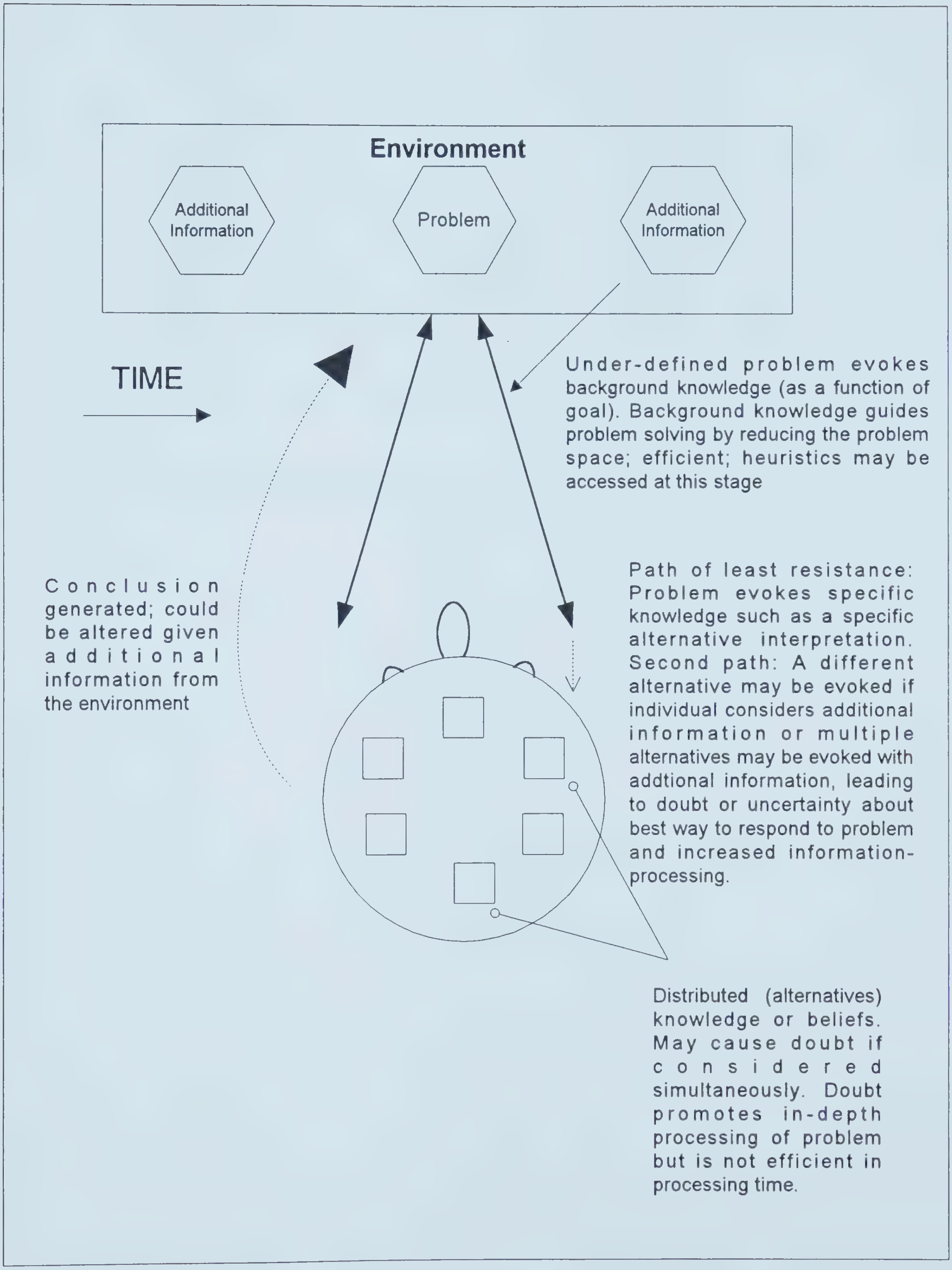


Figure 2.1. The Inductive-Coherence Model of Reasoning

(1)	(2)	(3)	(4)
M-P	P-M	M-P	P-M
<u>S-M</u>	<u>S-M</u>	<u>M-S</u>	<u>M-S</u>
S-P	S-P	S-P	S-P

Figure 2.2. The Four Traditional Figures of the Categorical Syllogism

First figure:	AAA	EAE	AII	EIO	AAI	EAO
Second figure:	EAE	AEE	EIO	AOO	EAO	AEO
Third figure:	AAI	IAI	AII	EAO	OAO	EIO
Fourth figure:	AAI	AEE	IAI	EAO	EIO	AEO

Figure 2.3. The 24 Valid Categorical Syllogisms

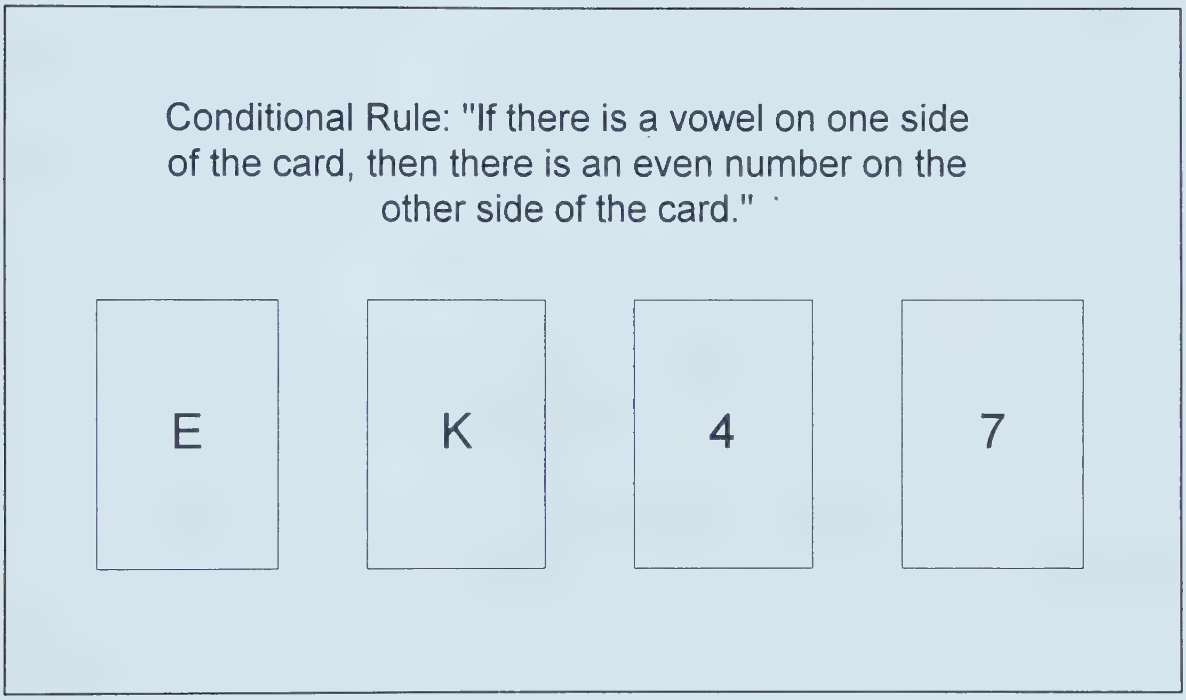


Figure 2.4. Wason's Card Selection Task

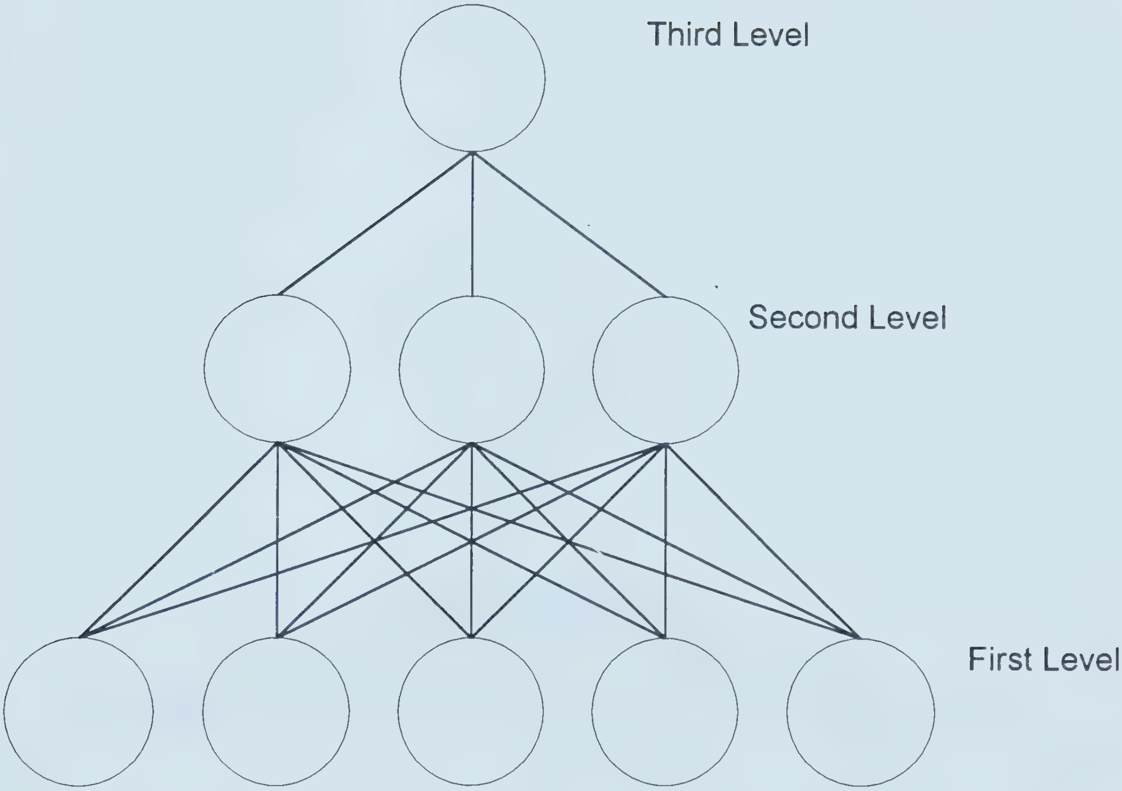


Figure 3.1. Illustration of a Typical PDP Network, Including Layer of Input Units, Hidden Units, and Output Units

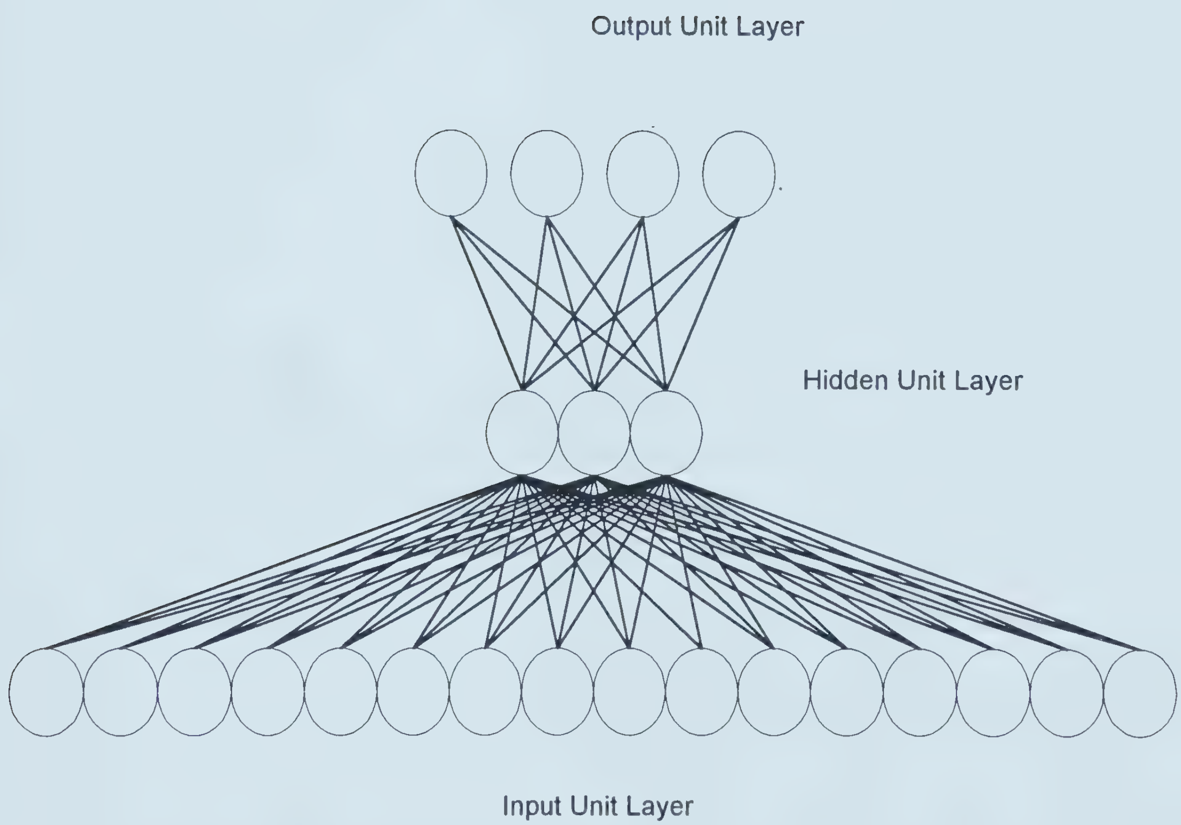


Figure 3.2. Illustration of the PDP Network Trained to Generate the Card Matching the Antecedent of the Conditional Rule (i.e., the "p" card).

Figure 3.3. Jittered Density Plots for Each of the Three Hidden Value Units Used In Network 1

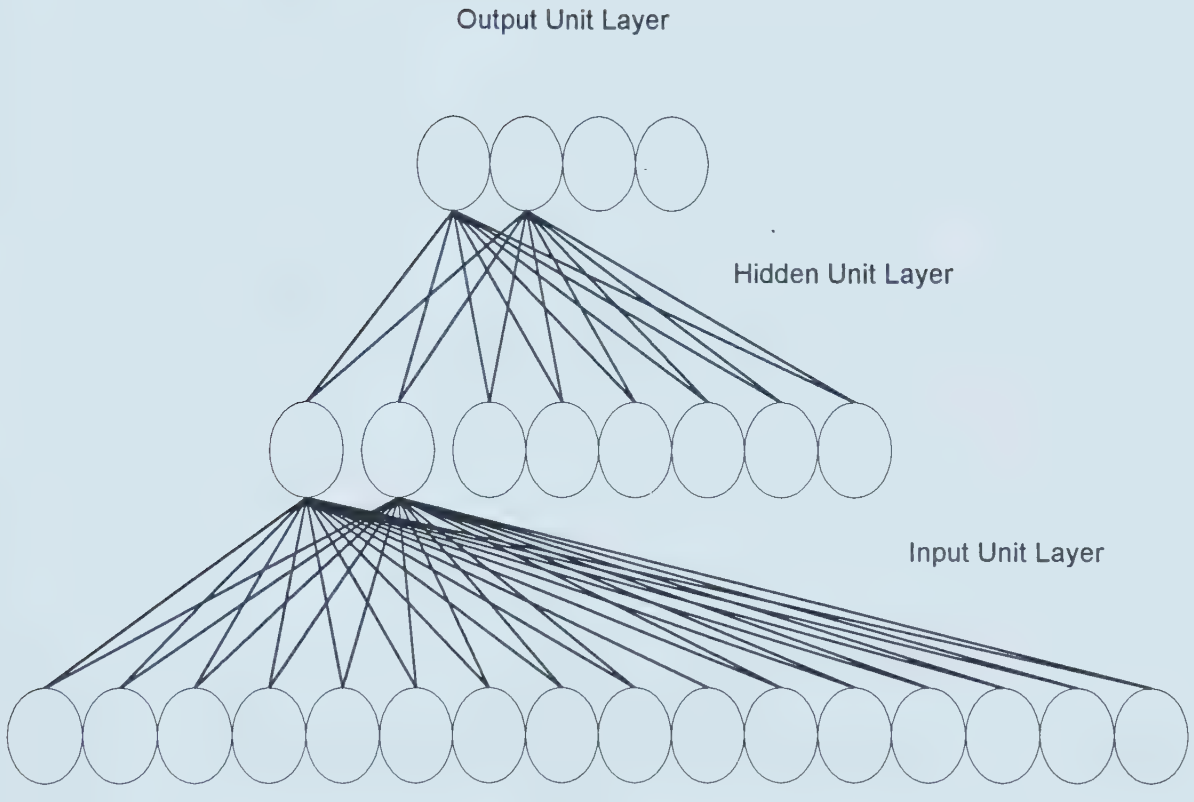


Figure 3.4. Illustration of the PDP Network Trained to Generate Both the Card Matching the Antecedent and the Card Negating the Consequent of the Conditional Rule (i.e., the "p" and "not-q" cards)

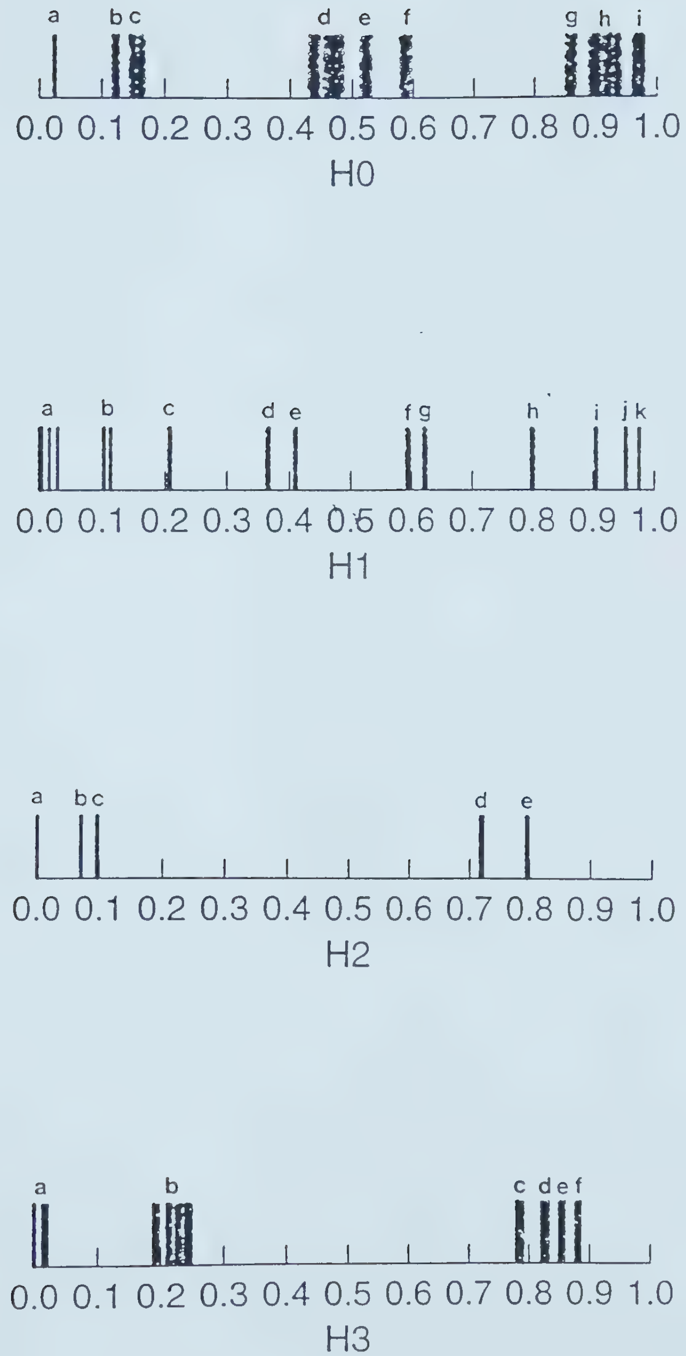


Figure 3.5. Jittered Density Plots for the First Set of 4 Hidden Value Units Used In Network 2

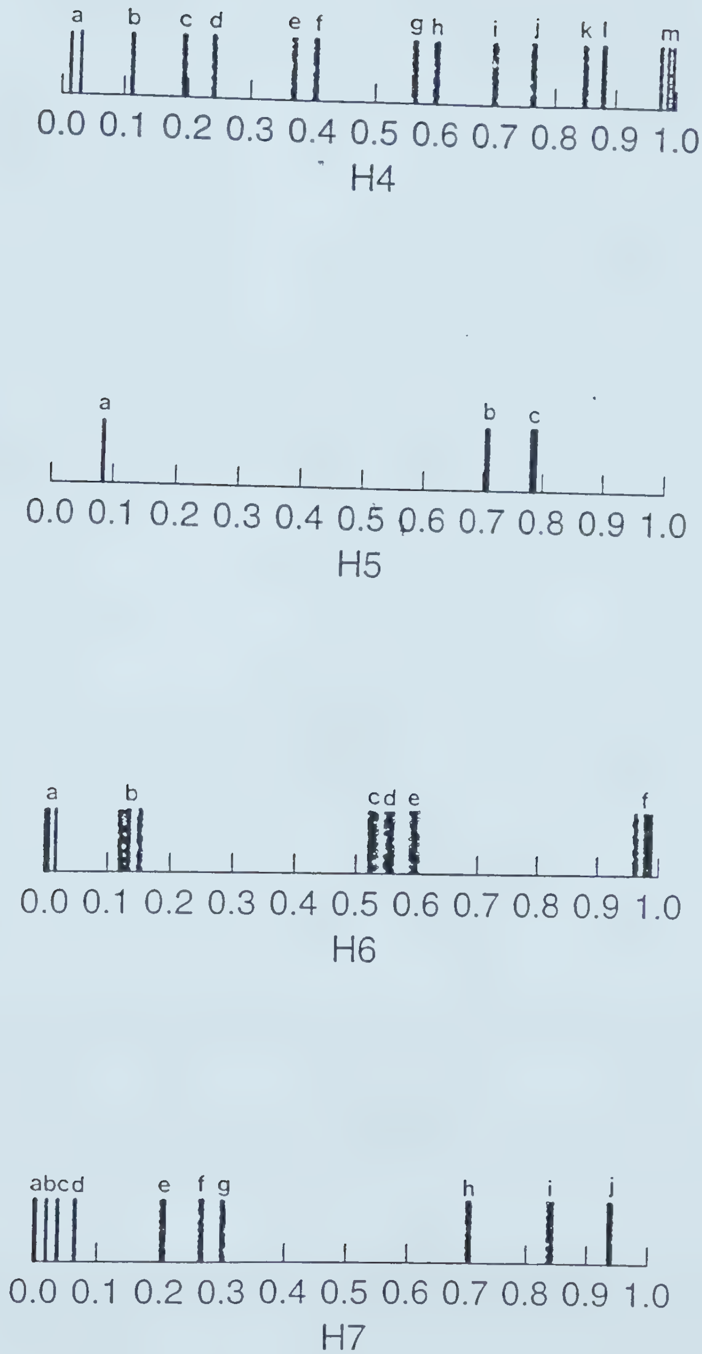


Figure 3.6. Jittered Density Plots for the Second Set of 4 Hidden Value Units Used In Network 2

Thematic Versions of Wason's Task

Rule Type

Card categories	Category	Instance
2	XX*	✓***
4	X**	X

*Condition violates purpose of task (i.e., cannot be experimentally tested)
**Condition has not been experimentally tested
***Condition has been experimentally tested

Abstract Versions of Wason's Task

Rule Type

Card categories	Category	Instance
2	XX*	X**
4	✓***	X

*Condition violates purpose of task (i.e., cannot be experimentally tested)
**Condition has not been experimentally tested
***Condition has been experimentally tested

Figure 4.1. The Experimental Studies of Thematic and Abstract Versions of Wason's Task That Have and Have Not Been Pursued, and That Cannot be Pursued

The Museum Caper: Batman to the Rescue

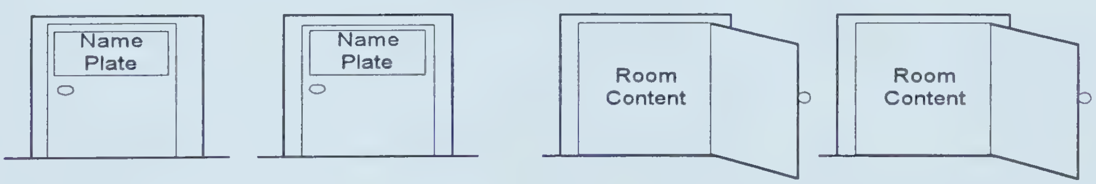
Two exhibitions, one entitled "Living Things" and a second entitled "Non-living Things," differing in familiarity and pleasantness have arrived at Gotham Museum.

Living Things Exhibition	Non-Living Things Exhibition
Land animals	Kitchen ware
Costumed professions	Bathroom ware
Aquatic animals	Office ware
Birds	Body ware

The Joker decides this is a perfect opportunity to ruin the Museum. The Museum's exhibition area is made up of sets of 4 rooms. A name plate on the door of each room identifies the room's content. The Joker decides to re-arrange all the name plates outside the rooms. Consequently, name plates DO NOT match their room contents. Before the Joker flees, he leaves a list of independent "rules," where each rule describes the relationship between a room's content and its name plate. For example,

If the name plate refers to a ?, then there is a ? inside the room.

Some of the rules are TRUE and some are FALSE. The Joker challenges administrators to figure out which rules are true and which ones are false by using sets of 2 doors and 2 rooms (shown below). For each rule, administrators must choose the door(s) that would have to be opened (to check its room content) and/or the door(s) that would have to be closed (to check its name plate) in order to figure out if the rule is true or false.



The Joker has also booby trapped the rooms to blow up if a door is needlessly opened and/or closed. Only the door(s) that are absolutely necessary to figure out the rule's truth or falsity must be opened and/or closed. Museum administrators call Batman for help. When Batman arrives he realizes that to preserve the Museum's safety he must play the Joker's diabolical game.

Pretend you are Batman. Your task is choose the door(s) that would have to be opened and/or closed to figure out if the Joker's rules are true or false. Please turn the page for an example of what you are to do.

Figure 4.2. Batman Task Used in Study 1

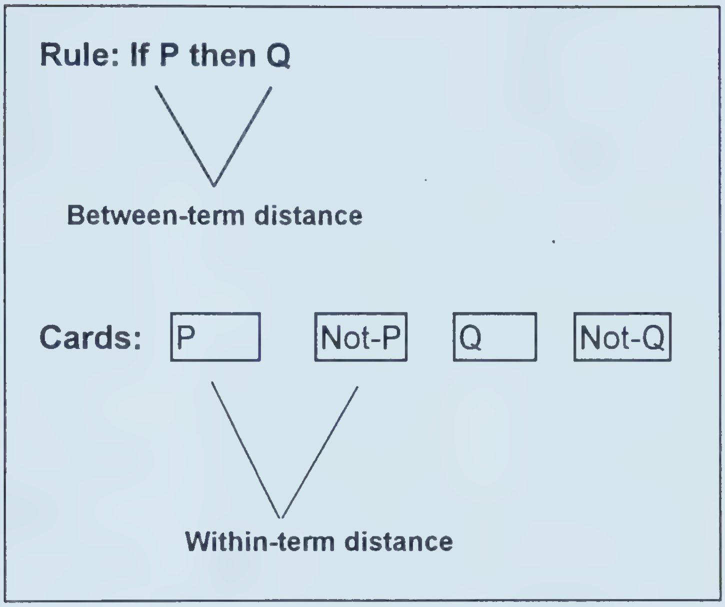


Figure 4.3. Illustration of Between-Term Distance and Within-term Distance

Higher-Level category:
Living Things

Categories:	Land Animals	Costumed Professions	Birds	Aquatic Animals
Instances:	beaver	acrobat	bluejay	bass
	deer	artist	robin	lobster
	dog	clown	trout	eagle
	squirrel	magician	whale	lark
	camel	bishop	catfish	cardinal
	goat	pastor	jellyfish	swallow
	mule	pope	sardine	pigeon
	sow	rabbi	carp	crow

Non-Living Things

Categories:	Kitchen Ware	Bathroom Ware	Office Ware	Body Ware
Instances:	mixer	bath	book	blouse
	pot	mirror	letters	dress
	refrigerator	shower	magazine	jacket
	stove	soap	money	pants
	burner	basin	closet	bow
	canopender	curler	typewriter	hood
	skillet	mat	scissors	shawl
	stool	tweezer	desk	cloak

Two categories considered low in distances would be taken from within the higher-level categories (e.g., land animals and aquatic animals); two categories considered high in distance would be taken from distinct higher-level categories (e.g., land animals and kitchen ware).

Figure 4.4. Illustration of Higher-Level Categories, Categories, and Instances Taken From Toglia and Battig (1978)

The Museum Caper: Batman to the Rescue

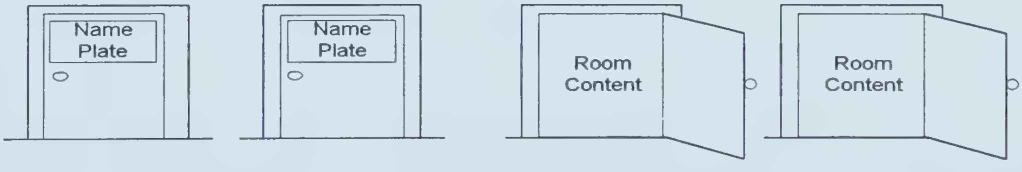
Two exhibitions, one entitled "Living Things" and a second entitled "Non-living Things," differing in familiarity and pleasantness have arrived at Gotham Museum.

Living Things Exhibition	Non-Living Things Exhibition
Land animals	Kitchen ware
Costumed professions	Bathroom ware
Aquatic animals	Office ware
Birds	Body ware

The Joker decides this is a perfect opportunity to ruin the Museum. The Museum's exhibition area is made up of sets of 4 rooms. A name plate on the door of each room identifies the room's content. The Joker decides to re-arrange all the name plates outside the rooms. Consequently, name plates DO NOT match their room contents. Before the Joker flees, he leaves a list of independent "rules," where each rule describes the relationship between a room's content and its name plate. For example,

If the name plate refers to a ?, then there is a ? inside the room.

Knowing the Joker's prankster style, all the rules are absolutely FALSE. The Joker challenges administrators to show that his rules are false with sets of 2 doors and 2 rooms (shown below). For each rule, administrators must choose the door(s) that would have to be opened (to check its room content) and/or the door(s) that would have to be closed (to check its name plate) in order to show that the rule is false.



The Joker has also booby trapped the rooms to blow up if a door is needlessly opened and/or closed. Only the door(s) that are absolutely necessary to show the rule's falsity must be opened and/or closed. Museum administrators call Batman for help. When Batman arrives he realizes that to preserve the Museum's safety he must play the Joker's diabolical game.

Pretend you are Batman. Your task is to choose the door(s) that would have to be opened and/or closed to show that the Joker's rules are false. Please turn the page for an example of what you are to do.

Figure 4.5. Determinate Batman Task Used in Study 2

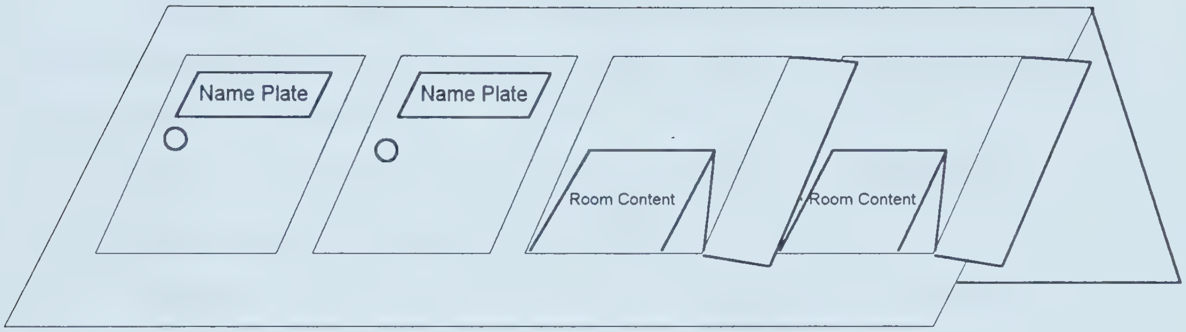


Figure 4.6. Miniature Museum Rooms Used to Test Participants Individually in Study 3

Bibliography

Ballard, D. (1986). Cortical structures and parallel processing: Structure and function. *The Behavioral and Brain Sciences, 9,* 67-120.

Bechtel, W., & Abrahamsen, A. (1991). *Connectionism and the mind.* Cambridge, MA: Blackwell.

Begg, I., & Denny, P. (1969). Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning errors. *Journal of Experimental Psychology, 81*, 351-354.

Begg, I., & Harris, G. (1982). On the interpretation of syllogisms. *Journal of verbal learning and verbal behavior, 21*, 595-620.

Bendixen, L.D., Dunkle, M.E., & Schraw, G. (1994). Epistemological beliefs and reflective judgement. *Psychological reports, 75*, 1595-1600.

Berkeley, I.S.N., Dawson, M.R.W., Medler, D.A., Schopflocher, D.P., & Hornsby, L. (1995). Density plots of hidden value unit activations reveal interpretable bands. *Connection Science, 7*, 167-186.

Bisanz, J., Bisanz, G.L., & Korpan, C.A. (1994). Inductive reasoning. In R.J. Sternberg (Ed.), *Thinking and problem solving: Handbook of perception and cognition (2nd ed.)*. CA: Academic Press.

Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review, 85*(1), 1-21.

Braine, M.D.S., & O'Brien, D.P. (1991). A theory of *If:* A lexical entry, reasoning program, and pragmatic principles. *Psychological review, 98*, 182-203.

Braine, M.D.S., & Rumain, B. (1983). Logical reasoning. In P.H. Mussen (Series Ed.) & J.H. Flavell & E.M. Markman (Vol. Eds.) *Handbook of child psychology: Vol. 3. Cognitive development* (4th ed., pp.263-340). New York: Wiley.

Byrne, R.M.J. (1989). Suppressing valid inferences with conditionals. *Cognition, 31*, 61-83.

Byrne, R.M.J. (1991). Can valid inferences be suppressed? *Cognition, 39*, 71-78.

Ceraso, J., & Provitera, A. (1971). Sources of error in syllogistic reasoning.

Cognitive psychology, 2, 400-410.

Chan, D., & Chua, F. (1994). Suppression of valid inferences: syntactic views, mental models, and relative salience. *Cognition, 53*, 217-238.

Cheng, P.W., & Holyoak, K.J. (1985). Pragmatic reasoning schemas. *Cognitive psychology, 17*, 391-416.

Cheng, P.W., & Holyoak, K.J. (1989). On the natural selection of reasoning theories. *Cognition, 33*, 285-313.

Cheng, P.W., & Nisbett, R.E. (1993). Pragmatic constraints on causal deduction. In R.E. Nisbett (Ed.), *Rules for Reasoning*. Hillsdale, NJ: Lawrence Erlbaum.

Cohen, L.J. (1981). Can human irrationality be experimentally demonstrated? *The behavioral and brain sciences, 4*, 317-370.

Collins, A., & Michalski, R. (1989). The logic of plausible reasoning: A core theory. *Cognitive science, 13*, 1-49.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how human reason? Studies with the Wason selection task. *Cognition, 31*, 187-276.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgement under uncertainty. *Cognition, 58*, 1-73.

Cox, J.R., & Griggs, R.A. (1982). The effects of experience on performance in Wason's selection task. *Memory & Cognition, 10*, 496-502.

Cummins, D.D. (1995). Naive theories and causal deduction. *Memory & Cognition, 23*, 646-658.

Cummins, D.D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition, 19*, 274-282.

Dawson, M.R.W. (1998). *Understanding cognitive science*. Blackwell.

Dawson, M.R.W., Medler, D.A., & Berkeley, I.S.N. (1997). PDP networks can provide models that are not mere implementations of classical theories. *Philosophical Psychology, 10*, 25-40.

Dawson, M.R.W., & Schopflocher, D.P. (1992). Modifying the generalized delta

rule to brain networks of non-monotonic processors for pattern classification. *Connection Science, 4*, 19-31.

Digdon, N.L. (1986). *Conditional and biconditional interpretations of if-then sentences: The role of content and context*. Unpublished Doctoral Dissertation.

Downing, C.J., Sternberg, R.J., & Ross, B.H. (1985). Multicausal inference: Evaluation of evidence in causally complex situations. *Journal of experimental psychology: General, 114*, 239-263.

Einhorn, H.J., & Hogarth, R.M. (1981). Behavioral decision theory: Processes of judgement and choice. *Annual review of psychology, 32*, 53-88.

Elio, R.(1997). What to believe when inferences are contradicted. The impact of knowledge type and inference rule. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum.

Elio, R., & Pelletier, F.J. (1994). The effect of syntactic form on simple belief revisions and updates. In *Proceedings of the 16th Annual Conference of the Cognitive Science Society*. Atlanta (pp.260-265).

Evans, J.St.B.T. (1989). *Bias in human reasoning: Causes and consequences*. Hillsdale: Lawrence Erlbaum.

Evans, J.St.B.T., Barston, J.L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition, 11*, 295-306.

Evans, J.St.B.T., & Lynch, J.S. (1973). Matching bias in the selection task. *British Journal of Psychology, 64*, 391-397.

Evans, St.B.T.J., Newstead, S.E., & Byrne, R.M. (1993). *Human reasoning: The psychology of deduction*. Hillsdale: Lawrence Erlbaum.

Evans, J.St.B.T., Over, D.E., & Manktelow, K.I. (1993). Reasoning, decision making, and rationality. *Cognition, 49*, 165-187.

Falmagne, R.J., & Gonsalves, J. (1995). Deductive inference. *Annual review of psychology, 46*, 525-559.

Fillenbaum, S. (1975). If: Some Uses. *Psychological Research, 37*, 245-260.

Fillenbaum, S. (1993). Deductive reasoning: What are taken to be the premises

and how are they interpreted? *Behavioral and brain sciences, 16*, 348-349.

Fodor, J.A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: The MIT Press.

Galotti, K.M. (1989). Approaches to studying formal and everyday reasoning. *Psychological Bulletin, 105*, 331-351.

Gärdenfors, P. (1992). Belief revision: An introduction. In P. Gärdenfors (Ed.), *Belief Revision*. Cambridge: Cambridge University Press.

Garner, W.R. (1962). *Uncertainty and structure as psychological concepts*. New York: Wliey.

Garnham, A., & Oakhill, J. (1994). *Thinking and reasoning*. Cambridge, MA: Blackwell.

Gigerenzer, G., & Hug, K. (1994). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43*, 127-171.

Gigerenzer, G., & Murray, D.J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.

Gilhoolly, K.J., Logie, R.H., Wetherick, N.E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory & Cognition, 21*, 115-124.

Ginsberg, M.L. (1986). Introduction. In M.L. Ginsberg (Ed.), *Readings in nonmonotonic reasoning*. Los Altos, CA: Morgan Kaufmann Publishers.

Goldstone, R.L. & Barsalou, L. (1998). Reuniting perception and conception. *Cognition, 65*, 231-62.

Grice, H.P. (1975). Logic and conversation. In P. Cole and J.L. Morgan (Eds.), *Syntax and semantics Volume 3: Speech acts* (pp.41-58). London: Academic Press.

Griggs, R.A. (1983). The role of problem content in the selection task. In J.St.B.T. Evans (Ed.), *Thinking and reasoning: Psychological approaches* (pp.16-43). London: Routledge and Kegan Paul.

Harman, G. (1986). *Change in view: Principles of reasoning*. Cambridge, MA: MIT Press.

Harman, G. (1995). Rationality. In E.E. Smith and D.N. Osherson (Eds.),

Thinking: An invitation to cognitive science (2^{nd}), Volume 3. Cambridge, MA: The MIT Press.

Henle, M. (1962). On the relation between logic and thinking. *Psychological Review, 69*, 366-378.

Hitch, G.J., & Baddeley, A.D. (1976). Verbal reasoning and working memory. *Quarterly Journal of Experimental Psychology, 28*, 603-621.

Holland, J.H., Holyoak, K.J., Nisbett, R.E., & Thagard, P.R. (1986). A framework for induction. *Induction: Processes of inference, learning, and discovery* (pp.1-28). Cambridge, MA: MIT Press.

Janis, I.L. & Frick, F. (1943). The relationship between attitudes toward conclusions and errors in judging logical validity of syllogisms. *Journal of experimental psychology, 33*, 73-77.

Johnson-Laird, P.N. (1994). Mental models and probabilistic thinking. *Cognition, 50*, 189-209.

Johnson-Laird, P.N. (1983). Inference and mental models. *Mental models. Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Johnson-Laird, P.N., & Bara, B.G. (1984). Syllogistic inference. *Cognition, 16*, 1-61.

Johnson-Laird, P.N., & Byrne, R.M.J. (1991). *Deduction*. Hillsdale: Lawrence Erlbaum.

Johnson-Laird, P.N., & Tagart, J. (1969). How implication is understood. *American Journal of Psychology, 82*, 367-373.

Kareev, Y., & Halberstadt, N. (1993). Evaluating negative tests and refutations in a rule discovery task. *The Quarterly Journal of Experimental Psychology, 46A*, 715-727.

Klayman, J., & Ha, Y-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological review, 94*, 211-228.

Lefford, A. (1946). The influence of emotional subject matter on logical reasoning. *The journal of general psychology, 34*, 127-151.

Legrenzi, P. (1969). Relations between language and reasoning about deductive rules. In G.B. Flores d'Arcais and W.J.M. Levelt (Eds.), *Advances in psycholinguistics*. London: North Holland.

Lepper, M.R., Ross, L., & Lau, R.R. (1986). Persistence of inaccurate beliefs about the self: Perseverance effects in the classroom. *Journal of personality and social psychology, 50*, 482-491.

Liberman, N., & Klar, Y. (1996). Hypothesis testing in Wason's selection task: Social exchange, cheating detection, or task understanding. *Cognition, 58*, 127-156.

Manktelow, K.I., & Evans, J.St.B.T. (1979). Facilitation of reasoning by realism: Effect or non-effect? *British journal of psychology, 70*, 477-488.

Manktelow, K.I., & Over, D.E. (1991). Social roles and utilities in reasoning with deontic conditionals. *Cognition, 39*, 85-105.

Markovits, H. (1984). Awareness of the 'possible' as a mediator of formal thinking in conditional reasoning problems. *British journal of psychology, 75*, 367-376.

Markovits, H. (1985). Incorrect conditional reasoning among adults: Competence or performance? *British Journal of Psychology, 76*, 241-247.

Markovits, H., & Guilaine, N. (1989). The belief-bias effect in the production and evaluation of logical conclusions. *Memory & Cognition, 17*, 11-17.

Moorehead, I.R., Haig, N.D., & Clement, R.A. (1989). An investigation of trained neural networks from a neurophysiological perspective. *Perception, 18*, 793-803.

Mynatt, C.R., Doherty, M.E., & Dragan, W. (1993). Information relevance, working memory, and the consideration of alternatives. *The Quarterly Journal of Experimental Psychology, 46A*, 759-778.

Neisser, U. (1982). *Memory Observed: Remembering in natural contexts.* San Francisco: W.H. Freeman.

Newell, A., & Simon, H.A. (1972). *Human problem solving*. New Jersey: Prentice-Hall.

Newstead, S.E. (1989). Interpretational errors in syllogistic reasoning. *Journal of memory and language, 28*, 78-91.

Newstead, S.E., & Evans, J.St.B.T. (1993), Mental models as an explanation of belief bias effects in syllogistic reasoning. *Cognition, 46*, 93-97.

Newstead, S.E., Pollard, P., Evans, J.St.B.T., & Allen, J.L. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition, 45*, 257-284.

Oakhill, J., & Garnham, A. (1993). On theories of belief bias in syllogistic reasoning. *Cognition, 46*, 87-92.

Oakhill, J.V., & Johnson-Laird, P.N. (1985). The effects of belief on the spontaneous production of syllogistic conclusions. *The quarterly journal of experimental psychology, 37A*, 553-569.

Oakhill, J., Johnson-Laird, P.N., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition, 31*, 117-140.

O'Brien, D.P. (1993). Mental logic and human irrationality: We can put a man on the moon, so why can't we solve those logical-reasoning problems? In K.I. Manktelow and D.E. Over (Eds.), *Rationality: Psychological and philosophical perspectives*. London: Routledge.

Oaksford, M., & Chater, N. (1993). Reasoning theories and bounded rationality. In K.I. Manktelow and D.E. Over (Eds.), *Rationality: Psychological and philosophical perspectives*. London: Routledge.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological review, 101*, 608-631.

Ormerod, T.C., Manktelow, K.I., & Jones, G.V. (1993). Reasoning with three types of conditional: Biases and mental models. *The Quarterly Journal of Experimental Psychology, 46A*, 653-677.

Osherson, D.N. (1990). Judgement. In D.N. Osherson & E.E. Smith (Eds.), *Thinking: An invitation to cognitive science* (pp.55-87). Cambridge, MA: MIT Press.

Petty, R.E., Priester, J.R., & Wegener, D.T. (1994). Cognitive processes in attitude change. In R.S.Jr., Wyer & T.K. Srull (Eds.), *Handbook of social cognition: Vol. 2. Applications* (2nd ed., pp.69-142). Hillsdale, NJ: LEA.

Piaget, J. (1977). *The development of thought: Equilibrium of cognitive*

structures. New York: Viking Press.

Politzer, G., & Braine, M.D.S. (1991). Responses to inconsistent premises cannot count as suppression of valid inferences. *Cognition, 38*, 103-108.

Pollard, P. (1982). Human reasoning: Some possible effects of availability. *Cognition, 12*, 65-96.

Pollard, P., & Evans, J.St.B.T. (1987). Content and context effects in reasoning. *American journal of psychology, 100*, 41-60.

Quine, W.V., & Ullian, J.S. (1978). *The web of belief (2nd edition)*. NY: Random House.

Rescher, N. (1980). *Induction*. Oxford: Blackwell.

Revlin, R., & V.O. Leirer. (1978). The effects of personal biases on syllogistic reasoning: rational decisions from personalized representations. In R. Revlin & R.E. Mayer (Eds.), *Human reasoning*. Washington, DC: Wiley.

Revlin, R., Leirer, V., Yopp, H., & Yopp, R. (1980). The belief-bias effect in formal reasoning: The influence of knowledge on logic. *Memory & Cognition, 8*, 584-592.

Rips, L.J. (1986). Mental muddles. In M. Brand and R.M. Harnish (Eds.), *The representation of knowledge and belief*. Tucson, AZ: The University of Arizona Press.

Rips, L.J. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.

Rips, L.J. (1995). Deduction and cognition. In E.E. Smith and D.N. Osherson (Eds.), *Thinking: An invitation to cognitive science (2nd edition), Vol. 3*. Cambridge, MA: MIT Press.

Rissland, E.L. (1986). Learning how to argue: Using hypotheticals. In J.L. Kolodner & C.K. Riesback (Eds.), *Experience, memory, and reasoning*. Hillsdale: Lawrence Erlbaum.

Roberts, M.J. (1993). Human reasoning: Deduction rules or mental models, or both? *The Quarterly Journal of Experimental Psychology, 46A*, 569-589.

Rumain, B., Connell, J., Braine, M.D.S. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: *If* is not the biconditional. *Developmental psychology, 19*, 471-481.

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature, 323*, 533-536.

Shoham, Y. (1990). Nonmonotonic reasoning and causation. *Cognitive science, 14*, 213-252.

Sloman, S.A. (1994). When explanations compete: the role of explanatory coherence on judgements of likelihood. *Cognition, 52*, 1-21.

Sloman, S.A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. Cognitive Psychology. *Cognition, 35*, 1-33.

Sloman, S.A., & Rips, L.J. (1998). Similarity as an explanatory construct. *Cognition, 65*, 87-101.

Staudenmayer, H. (1975). Understanding conditional reasoning with meaningful propositions. In R.J. Falmagne (Ed.), *Reasoning: Representation and process in children and adults*. Hillsdale: Lawrence Erlbaum.

Stevenson, R.J., & Over, D.E. (1995). Deduction from uncertain premises. *The Quarterly Journal of Experimental Psychology, 48A*, 613-643.

Swain, M. (1979). Justification and belief. In G.S. Pappas (Ed.), *Justification and knowledge*. Boston: Reidel.

Taplin, J.E. (1971). Reasoning with conditional sentences. *Journal of verbal learning and verbal behavior, 10*, 219-225.

Taplin, J.E., & Staudenmayer, H. (1973). Interpretation of abstract conditional sentences in deductive reasoning. *Journal of verbal learning and verbal behavior, 12*, 530-542.

Thagard, P. (1989). Explanatory coherence. *Behavioral and brain sciences, 12*, 435-502.

Toglia, M. P. & Battig, W.F. (1978). *Handbook of Semantic Word Norms*. New Jersey: Erlbaum.

Toms, M., Morris, N., & Ward, D. (1993). Working memory and conditional reasoning. *The Quarterly Journal of Experimental Psychology, 46A*, 679-699.

Tourangeau, R., & Sternberg, R. (1981). Aptness in metaphor. *Cognitive*

Psychology, 13, 27-55.

Tourangeau, R., & Sternberg, R. (1982). Understanding and appreciating metaphors. *Cognition, 11*, 203-244.

Tversky, A., & Kahneman, D. (1982). Availability: A heuristic for judging frequency and probability. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Wason, P.C. (1966). Reasoning. In B.M. Foss (Ed.), *New Horizons in Psychology*. Penguin.

Wason, P.C. (1977). Self-contradictions. In P.N. Johnson-Laird and P.C. Wason (Eds.), *Thinking: Readings in cognitive science*. Cambridge: Cambridge University Press.

Wason, P.C. (1983). Realism and rationality in the selection task. In J.St.B.T. Evans (Ed.), *Thinking and reasoning: Psychological approaches*. London: Routledge and Kegan Paul.

Wason, P.C., & Johnson-Laird, P.N. (1972). *Psychology of reasoning: Structure and content*. London: Batsford. Cambridge, MA: Harvard University Press.

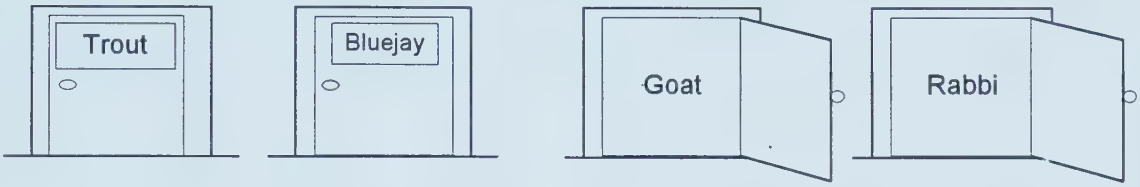
Wetherick, N.E., & Gilhooly, K.J. (1990). Syllogistic reasoning. Effects of premise order. In K.J. Gilhooly, M.T.G. Keane, R.H. Logie, & G. Erdos (Eds.), *Lines of thinking: Reflections on the psychology of thought (vol.1)*.Chichester: Wiley.

Woodworth, R.S., & Sells, S.B. (1935). An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology, 18*, 451-460.

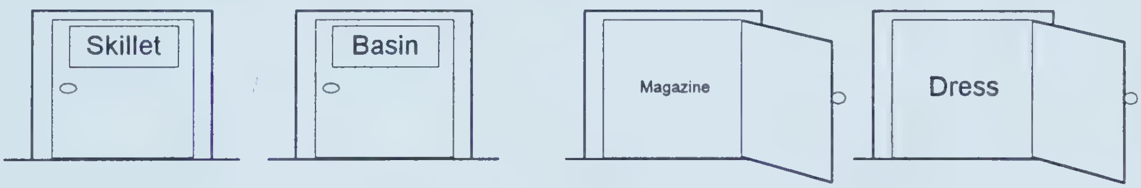
Appendix 1 (Study 1). CONDITION 1 Batman Task

For each rule, Batman must choose the door(s) that would have to be opened and/or the door(s) that would have tobe closed in order to figure out if the rule is true or false. Please consider each rule independently (rules are not related to each other).
Please choose the door(s) in order of importance by placing a "1" over the first door you would choose etc. Remember, some of the Joker's rules are true and some are false.

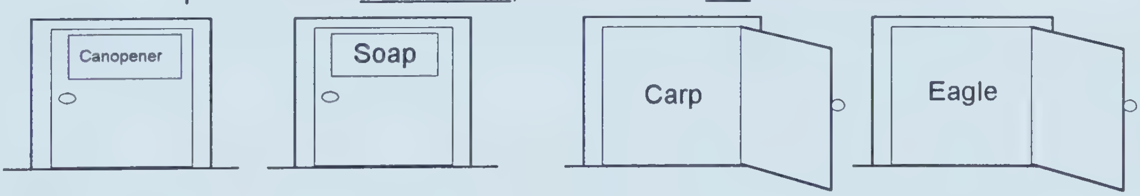
1. If the name plate refers to a bird, then there is a costumed profession inside the room.



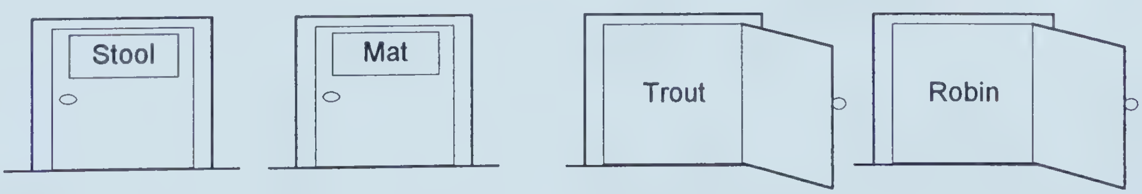
2. If there is office ware inside the room, then the name plate refers to bathroom ware.



3. If the name plate refers to kitchen ware, then there is a bird inside the room.

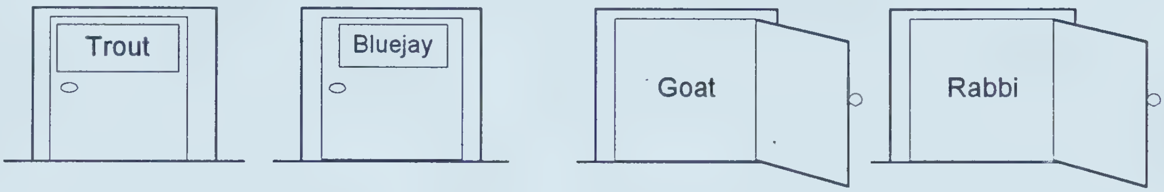


4. If there is a bird inside the room, then the name plate refers to bathroom ware.

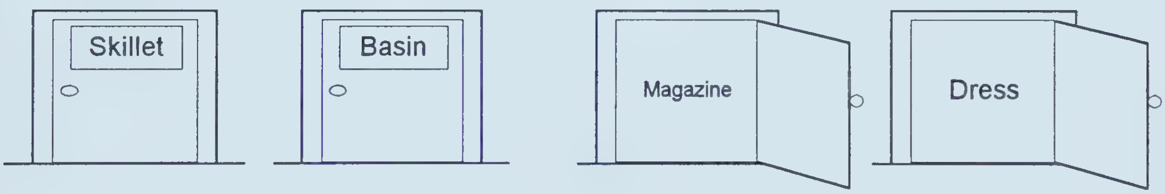


For each rule, Batman must choose the door(s) that would have to be opened and/or the door(s) that would have to be opened and/or closed in order to figure out if the rule is true or false. Please consider each rule independently (rules are not related to each other).
Please choose the door(s) in order of importance by placing a "1" over the first door you would choose etc. Remember, some of the Joker's rules are true and some are false.

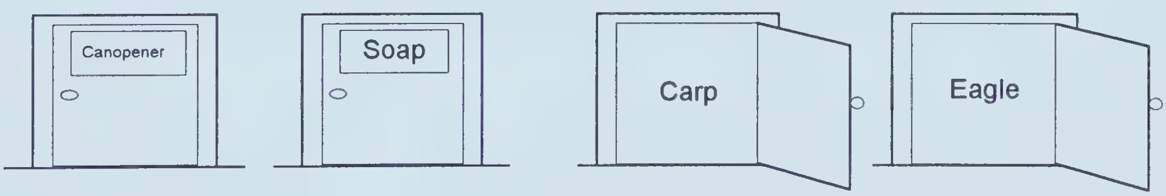
1. If the name plate refers to a bluejay, then there is a rabbi inside the room.



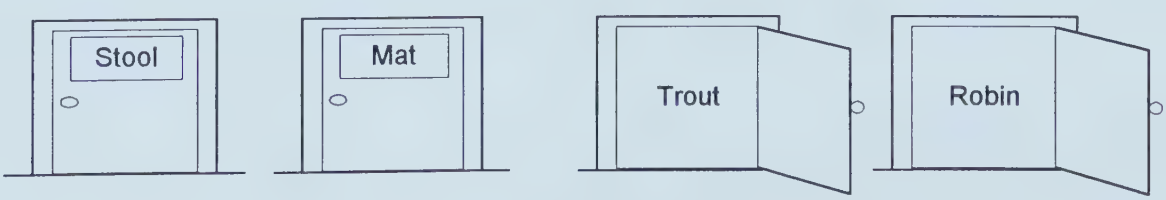
2. If there is a magazine inside the room, then the name plate refers to a basin.



3. If the name plate refers to a canopener, then there is an eagle inside the room.



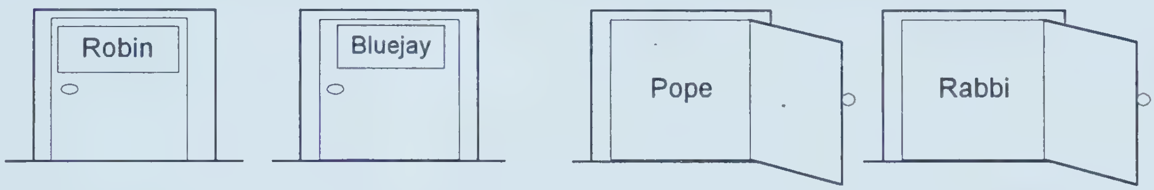
4. If there is a robin inside the room, then the name plate refers to a mat.



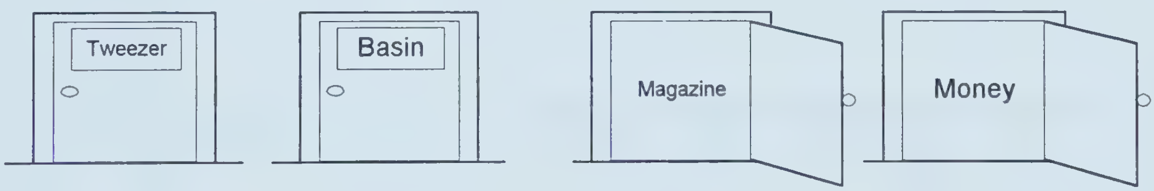
Appendix 1 (Study 1). CONDITION 3 Batman Task

For each rule, Batman must choose the door(s) that would have to be opened and/or the door(s) that would have to be closed in order to figure out if the rule is true or false. Please consider each rule independently (rules are not related to each other).
<u>Please choose the door(s) in order of importance by placing a "1" over the first door you would choose etc</u>. Remember, some of the Joker's rules are true and some are false.

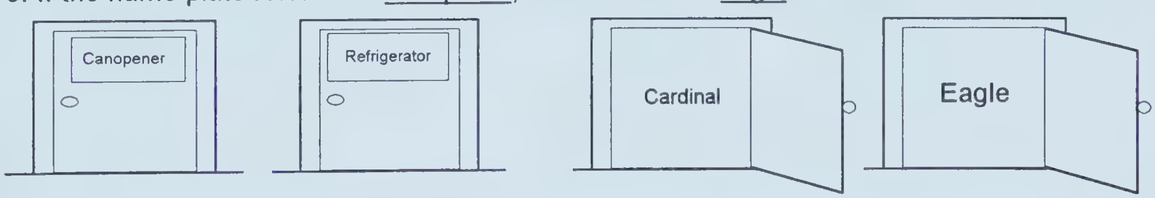
1. If the name plate refers to a <u>bluejay</u>, then there is a <u>rabbi</u> inside the room.



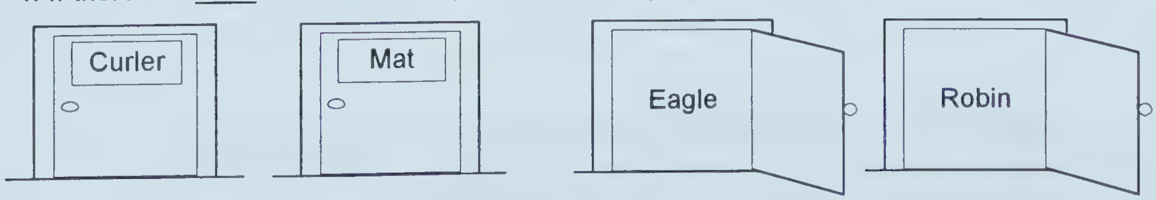
2. If there is a <u>magazine</u> inside the room, then the name plate refers to a <u>basin</u>.



3. If the name plate refers to a <u>canopener</u>, then there is an <u>eagle</u> inside the room.



4. If there is a <u>robin</u> inside the room, then the name plate refers to a <u>mat</u>.



Appendix 1 (Study 1). CONDITION 1 Wason's Task

Imagine you are presented with a rule describing a relationship between letters and numbers. Imagine also four cards lying flat on a table (you can see only one side of each card), where each card has a letter on one side and a number on the other side.
Please read over each of the rules below and select the card(s) that would have to be turned over in order to test the truth or falsity of the rule. Please select the card(s) in order of importance by placing a "1" over the first card you would select etc.

1. If there is an odd number on one side, then there is a consonant on the other side.

A	D	3	6

2. If there is a vowel on one side, then there is an odd number on the other side.

G	E	7	4

3. If there is an even number on one side, then there is a consonant on the other side.

E	G	4	5

4. If there is a consonant on one side, then there is an even number on the other side.

C	E	4	3

5. If there is an odd number on one side, then there is a vowel on the other side.

G	A	6	7

6. If there is a vowel on one side, then there is an even number on the other side.

A	B	2	3

Appendix 1 (study 1). CONDITION 2 Wason's task

Imagine you are presented with a rule describing a relationship between letters and numbers. Imagine also four cards lying flat on a table (you can see only one side of each card), where each card has a letter on one side and a number on the other side.
Please read over each of the rules below and select the card(s) that would have to be turned over in order to test the truth or falsity of the rule. Please select the card(s) in order of importance by placing a "1" over the first card you would select etc.

1. If there is a 3 on one side, then there is a D on the other side.

A	D	3	6

2. If there is an E on one side, then there is a 7 on the other side.

G	E	7	4

3. If there is a 4 on one side, then there is an G on the other side.

E	G	4	5

4. If there is a C on one side, then there is a 4 on the other side.

C	E	4	3

5. If there is a 7 on one side, then there is an A on the other side.

G	A	6	7

6. If there is an A on one side, then there is a 2 on the other side.

A	B	2	3

Imagine you are presented with a rule describing a relationship between letters and numbers. Imagine also four cards lying flat on a table (you can see only one side of each card), where each card has a letter on one side and a number on the other side.
Please read over each of the rules below and select the card(s) that would have to be turned over in order to test the truth or falsity of the rule. Please select the card(s) in order of importance by placing a "1" over the first card you would select etc.

1. If there is a 3 on one side, then there is a D on the other side.

G	D	3	5

2. If there is an E on one side, then there is a 7 on the other side.

A	E	7	5

3. If there is a 4 on one side, then there is an G on the other side.

D	G	4	6

4. If there is a C on one side, then there is a 4 on the other side.

C	B	4	2

5. If there is a 7 on one side, then there is an A on the other side.

E	A	5	7

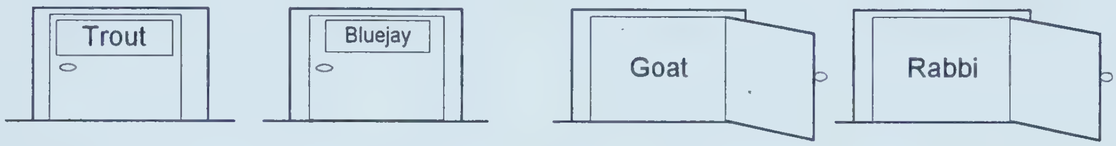
6. If there is an A on one side, then there is a 2 on the other side.

A	E	2	4

Appendix 2 (Study 2). CONDITION 1 Determinate Batman task

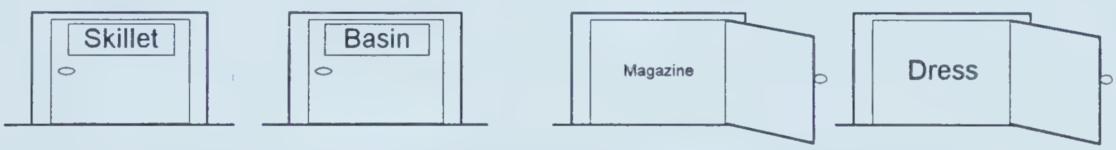
1. If the name plate refers to a bird, then there is a costumed profession inside the room.

Which door(s) would you choose in order to show that this rule is false?



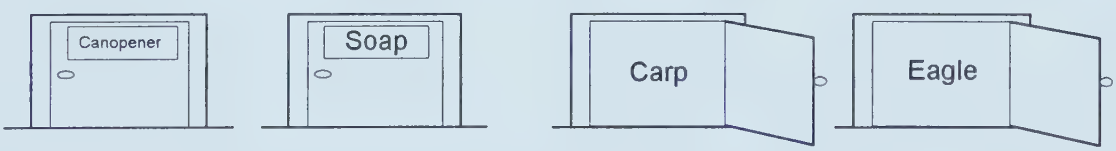
2. If there is office ware inside the room, then the name plate refers to bathroom ware.

Which door(s) would you choose in order to show that this rule is false?



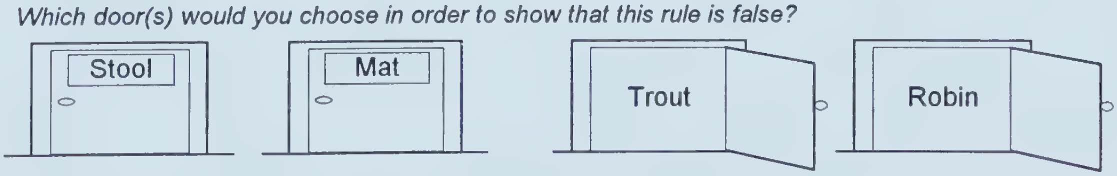
3. If the name plate refers to kitchen ware, then there is a bird inside the room.

Which door(s) would you choose in order to show that this rule is false?



4. If there is a bird inside the room, then the name plate refers to bathroom ware.

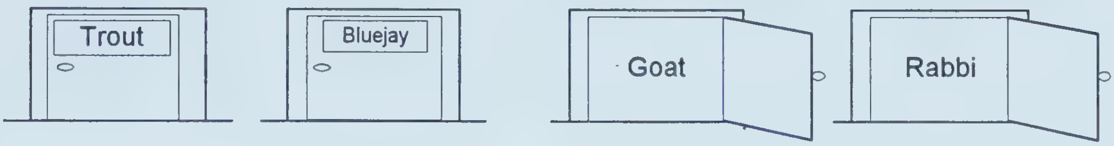
Which door(s) would you choose in order to show that this rule is false?



Appendix 2 (Study 2). CONDITION 2 Determinate Batman Task

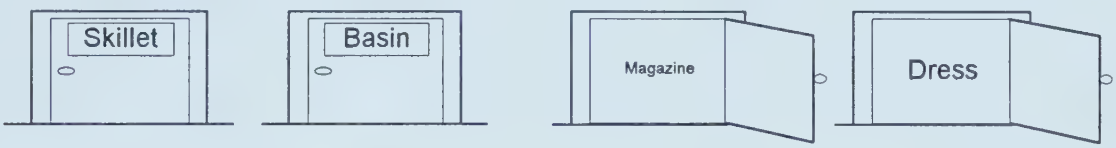
1. If the name plate refers to a bluejay, then there is a rabbi inside the room.

Which door(s) would you choose in order to show that this rule is false?



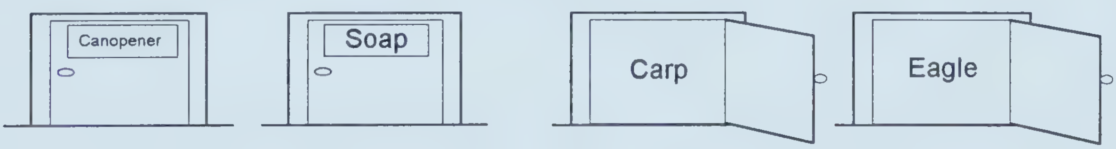
2. If there is a magazine inside the room, then the name plate refers to a basin.

Which door(s) would you choose in order to show that this rule is false?



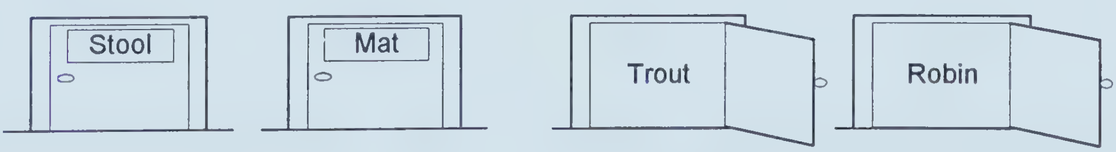
3. If the name plate refers to a canopener, then there is an eagle inside the room.

Which door(s) would you choose in order to show that this rule is false?



4. If there is a robin inside the room, then the name plate refers to a mat.

Which door(s) would you choose in order to show that this rule is false?



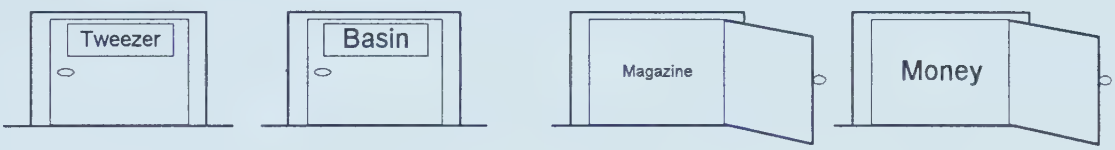
Appendix 2 (Study 2). CONDITION 3 Determinate Batman Task

1. If the name plate refers to a bluejay, then there is a rabbi inside the room.

Which door(s) would you choose in order to show that this rule is false?

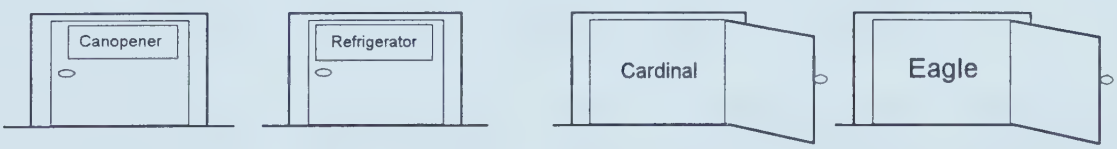
2. If there is a magazine inside the room, then the name plate refers to a basin.

Which door(s) would you choose in order to show that this rule is false?



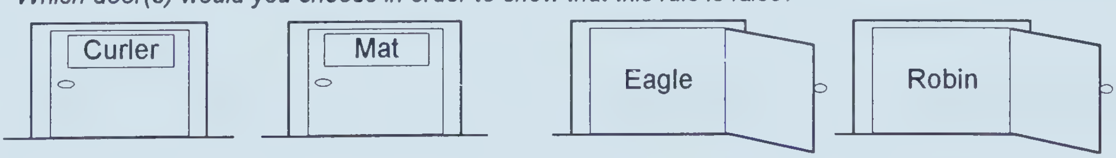
3. If the name plate refers to a canopener, then there is an eagle inside the room.

Which door(s) would you choose in order to show that this rule is false?



4. If there is a robin inside the room, then the name plate refers to a mat.

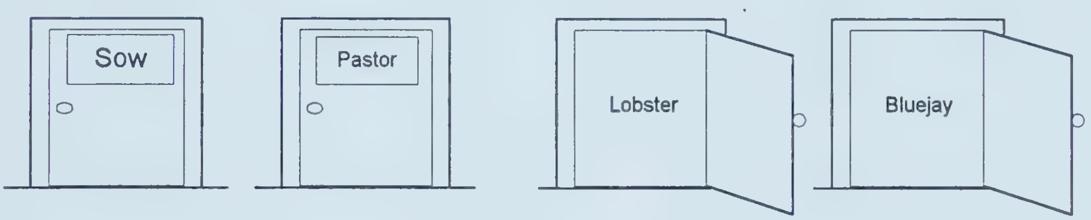
Which door(s) would you choose in order to show that this rule is false?



Appendix 3 (Study 3). CONDITION 1 Indeterminate & Determinate Batman Task

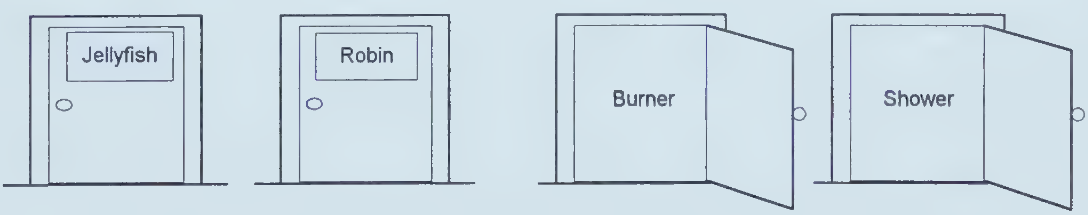
13. If the name plate refers to land animal, then there is a bird inside the room.

Which door(s) would you choose in order to figure out if this rule is true or false?
Which door(s) would you choose in order to show that this rule is false?



19. If the name plate refers to bird, then there is kitchen ware inside the room.

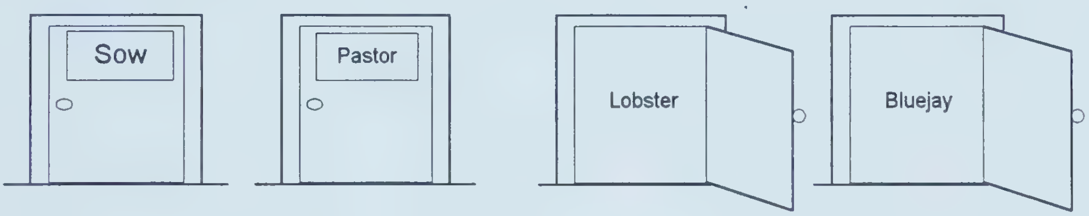
Which door(s) would you choose in order to figure out if this rule is true or false?
Which door(s) would you choose in order to show that this rule is false?



Appendix 3 (Study 3). CONDITION 2 Indeterminate & Determinate Batman Task

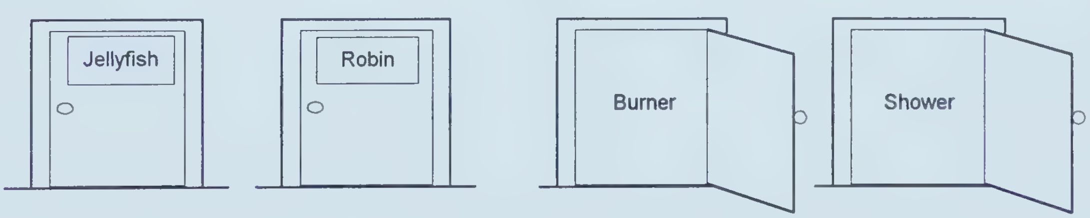
13. If the name plate refers to sow, then there is a bluejay inside the room.

Which door(s) would you choose in order to figure out if this rule is true or false?
Which door(s) would you choose in order to show that this rule is false?



19. If the name plate refers to robin, then there is burner inside the room.

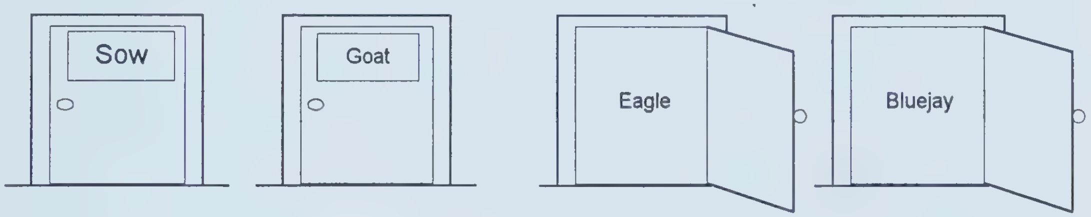
Which door(s) would you choose in order to figure out if this rule is true or false?
Which door(s) would you choose in order to show that this rule is false?



Appendix 3 (Study 3). CONDITION 3 Indeterminate & Determinate Batman Task

13. If the name plate refers to sow, then there is a bluejay inside the room.

Which door(s) would you choose in order to figure out if this rule is true or false?
Which door(s) would you choose in order to show that this rule is false?



19. If the name plate refers to robin, then there is burner inside the room.

Which door(s) would you choose in order to figure out if this rule is true or false?
Which door(s) would you choose in order to show that this rule is false?